Identification of Candidate Genes Controlling Porcine Female Reproductive Traits

Annemarie H. King

Thesis submitted for the degree

of

Doctor of Philosophy

The University of Edinburgh

2003

DECLARATION

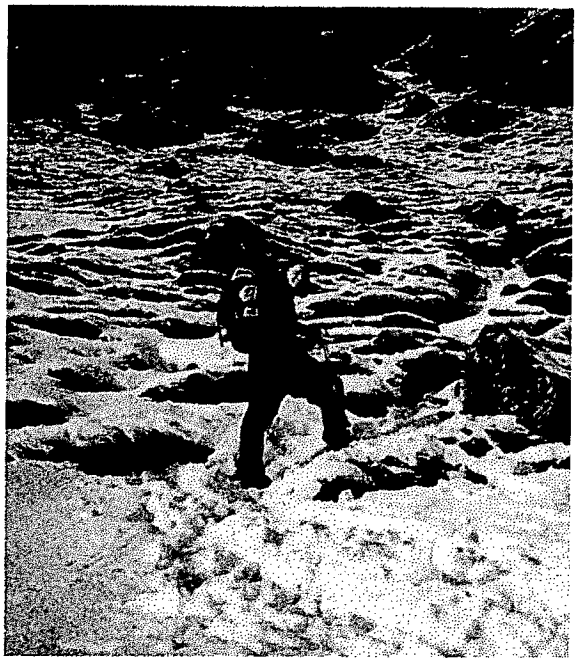
I declare that this thesis is the product of my own efforts and has not been submitted in any previous application for a degree. The research on which it is based is my own, except where stated in the text.

Annemarie King

Mount Improbable rears up from the plain, lofting its peaks dizzily to the rarefied sky. The towering, vertical cliffs of Mount Improbable can never, it seems, be climbed. Our mountaineers are too ambitious. So intent are they on the perpendicular drama of the cliffs, they do not think to look round the other side of the mountain.

There they would find not vertical cliffs and echoing canyons but gently inclined grassy meadows, graded steadily and easily towards the distant uplands. Occasionally a small, rocky crag punctuates the gradual ascent, but you can usually find a detour that is not too steep. The sheer height of the peak doesn't matter, so long as you don't try to scale it in a single bound. Locate the mildly sloping path and the ascent is only as formidable as the next step.

by Richard Dawkins, 1996. Climbing mount improbable.
Excerpts from Chapter 3 - Message from the Mountain



"The more I learn, the more I learn how little I know."
Socrates (469-399 B.C.E.)

ABSTRACT

Breeding programs have been used to improve livestock performance by selecting for traits of economic importance. However, as traits for prolificacy tend to have a low heritability, are sex limited and are expressed only in mature animals, improvement has been slow. The chromosomal locations that control these traits for which there is a continuum between high and low performing animals, are termed Quantitative Trait Loci (QTL). Three-generation pedigrees, in which the founders were Meishan and Large White purebred pigs, were used to identify QTL for female reproductive performance. The Chinese Meishan is one of the most prolific pig breeds known, farrowing up to five more piglets per litter than the European commercial Large White breed. However, the Meishan is not commercially viable in Europe due to its poor growth rate and high carcass fat content. Therefore, including the beneficial alleles from the Meishan into the Large White breed is of commercial relevance.

QTL for ovulation rate, teat number, age at puberty and uterine capacity have been mapped to chromosome 8 (SSC8) in earlier studies. The aim of this study was to focus specifically on SSC8 and to identify and test candidate genes for the QTL. The genotypes of twenty markers on SSC8 were combined with data collected on the reproductive performance of 220 F2 females. QTL for the related traits of litter size and prenatal survival were identified at the distal end of the q arm. The beneficial alleles at these QTL seem to be from the Meishan and that is consistent with the fact that this breed delivers larger litter sizes through higher levels of prenatal survival at a given ovulation rate.

The QTL for prenatal survival was defined as a region of about 30 cM and therefore contains many positional candidate genes, but only a few of these have been mapped in pigs. Human chromosome 4 (HSA4) shares extensive homology to SSC8. A gene map of SSC8 was developed by radiation hybrid mapping in order to align the human genome sequence of HSA4 with the QTL and thus identify comparative positional candidate genes. The resulting comparative map revealed extensive conservation of synteny and gene order. One positional candidate gene, secreted phosphoprotein 1 (*SPP1*), is also a physiological candidate gene as the protein is involved in porcine embryo implantation and maintenance of pregnancy.

A copy of *SPP1* (~10 kb) from both Meishan and Large White origin was sequenced to identify candidate causal variation. A total of 97 variants, including single nucleotide polymorphisms (SNPs) and sequence insertions and deletions were found. Three of the SNPs would result in non-synonymous amino acid changes. Expression of *SPP1* is controlled mainly by posttranslational modification and for one of these SNPs the Meishan allele encodes a serine, which would be phosphorylated and the Large White allele encodes a proline, which would not. To test whether the variation at this particular locus resulted in differences in litter size, the SNP was genotyped over 2974 pigs of varying breed origin. No association was identified between genotypes at this locus and litter size.

Evidence for genetic variation associated with litter size in or close to the *SPP1* gene comes from two other independent studies in which *SPP1* was used as the genetic marker. Future work could therefore involve typing further candidate SNPs in the *SPP1* gene to identify whether there is a causative mutation in this gene. Furthermore, additional positional candidate genes could be tested for causal variation.

ACKNOWLEDGEMENTS

I would like to thank my principal supervisor Alan Archibald for his continual encouragement and support and for his guidance through the more difficult genetics concepts, which I had to deal with. In addition I'd like to acknowledge the contribution made by my academic supervisor Peter Keightley and thank him for useful discussions with him. A debt of gratitude is owed to Chris Haley for his advice on the QTL analysis and for helping me to understand such a challenging subject!

Sygen has proved itself to be an exceptional sponsor company and my time with them has broadened my general skills, given me the opportunity to meet with other scientists and offered excellent work experiences, which I can carry forward into my scientific career. Carole Sargent, my supervisor from Sygen, has given me invaluable guidance during my five months spent in her laboratory in Cambridge, particularly in the areas of SNP detection and the use of microarrays. Her moral support and technical advice extended beyond Cambridge to Edinburgh throughout the three years. I'd particularly like to thank everyone I've worked with at Sygen, especially Graham Plastow.

There have been many collaborators that have made this project possible. By no means least I'd like to thank those at the Roslin Institute who have played a vital role in the PiGMaP reproductive QTL study over the last 15 years. These include specifically Heather Finlayson who had completed the mammoth task of preparing and organising all the DNA samples before I arrived; Andy Law, Richard Papworth, Geoff Lee, the staff at Dryden Farm and of course the pigs themselves!

It is important here to offer special thanks to everyone in the PiGMaP laboratory here at Roslin for showing me the ropes and making the three years so enjoyable. They gave me vital encouragement to believe that putting my social life on hold to write this thesis would be worthwhile in the end!! I'm very grateful for the opportunity to study in Edinburgh, not only to live in such a beautiful city, but also for the chance to explore and experience the most breath-taking mountains in Britain. The Ceilidh dancing, kilted men, Scottish ale and late licensing laws were added bonuses!!

My family and friends were always there for me and I would especially like to thank Anita and Nia for feeding me copious amounts of chocolate during my time off due to illness early on in the research. They helped me to build up the strength again to return to work. I would also like to thank my Mum for her continual support and for making sure I didn't work too hard and remembered to have fun. I would especially like to thank my fiancé Martin for all his understanding and support during the difficult times. We've been putting off planning our wedding until our respective projects were complete, now we can organise this new venture!

I'd like to dedicate this thesis to my Father. Unfortunately, due to his premature death, we sadly never had the opportunity to get to know one another.

PUBLICATIONS

The following publications have resulted as a direct outcome of the research described in this thesis:

- King,A.H., Jiang,Z.H., Gibson,J.P., Haley,C.S. and Archibald,A.L. Mapping of quantitative trait loci affecting female reproductive traits on porcine chromosome 8. *Biology of Reproduction*, Published In Press.

- King,A.H., Jiang,Z.H., Rohrer,G.A., Gibson,J.P., Waddington,D. and Archibald,A.L. Use of radiation hybrid mapping to locate candidate genes for female reproductive traits on porcine chromosome 8. Proceedings of the XXVIIIth International Conference on Animal Genetics, Göttingen, Germany, August 11-15, 2002. (available online at http://www.isag.org.uk).

- Annemarie H. King, Chris S. Haley and Alan L. Archibald. Quantitative trait loci (QTL) for female reproductive traits on porcine chromosome 8. *Genetical Research*, Oct 2001; 78(2):191-206

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ANOVA – analysis of variance
BAC – bacterial artificial chromosome
BLUP – best linear unbiased prediction
BMP – bone morphogenetic protein
bp - base pairs
BSA – bovine chromosome
cDNA – complementary deoxyribonucleic acid
CDS – coding sequence
cM – centiMorgan
cR – centiRays
DNA – deoxyribonucleic acid
dNTP – deoxynucleoside triphosphate
EST – expressed sequence tag
HSA – human chromosome
Kb - kilo base pairs
QTL - quantitative trait locus
LD – linkage disequilibrium
LW – Large White breed
MAS – marker assisted selection
MAI – marker assisted introgression
MMU – mouse chromosome
MS – Meishan breed
PCR – polymerase chain reaction
RH – radiation hybrid
RFLP – restriction fragment length polymorphism
SD – standard deviation
SEM – standard error of the mean
SINE - short interspersed element (mammalian-wide interspersed repeat)
SNP – single nucleotide polymorphism
SPP1 – secreted phosphoprotein 1
SSC - porcine chromosome
STS – sequence tagged site
U – units
UTR – untranslated region

Chapter One

1. INTRODUCTION

1.1. Genome mapping projects

The characteristics and performance of humans and animals are controlled by their genetic make-up, inherited from their parents and determined by external environmental influences. The proportion of the observed phenotypic variance, which is genetic in origin, rather than environmental, is defined as 'heritability'. By identifying genetic variation across a species genome, it is possible to gain a better understanding of the contributions of nature and nurture to physical and behavioural traits (Chakravarti, 2001). Indeed the last few decades have seen a rapid advance in our understanding of the highly complex genetic control of many traits in various species through whole genome mapping, sequencing and gene-expression studies.

To date the whole genomes of many viruses, naturally occurring plasmids, organelles, eubacteria, yeast, plants and animals have been sequenced. Specific examples include the completion of the genome sequence for the eubacteria *Escherichia coli* in 1997 (Blattner, 1997), for the worm *Caenorhabditis elegans* in 1998 (The Arabidopsis Genome Inititive, 1998), the fruit fly *Drosophila melanogaster* and the flowering plant *Arabidopsis thaliana* in 2000 (Adams *et al.*, 2000 and The *C. elegans* Sequencing Consortium, 2000), the human in 2001 (Lander et al., 2001) and finally the rice *Oryza sativa* and the mouse in 2002 (Yu *et al.*, 2002 and (Waterston et al., 2002).

1.1.1. Humans

The goal of the human genome project has been to map and sequence the whole genome and it has been said to be one of the most important projects in biology and the biomedical sciences that will permanently change both of these fields (Collins et al., 1998). The last century of scientific progress began with the discovery of Mendel's laws of the genetics of inheritance in the early 1900s and the current century started with the completion of the draft sequence of the whole genome in 2001.

Early geneticists realised that chromosomes mediate the transmission of Mendelian inheritance but they found it hard to comprehend how the featureless material of "chromatin" could control the development of a unique human being (Olson, 2002). Then in 1953 James Watson and Francis Crick described the double helix structure of deoxyribonucelic acid (DNA), which explained its ability to self-replicate and to direct the synthesis of proteins via the four nitrogenous organic bases (adenine, cytosine, guanine and thymine) (Watson and Crick, 1953).

Soon afterwards the mechanisms of transcription of genes via messenger ribonucleic acid (mRNA), the transfer of this mRNA from the nucleus to the cytoplasm of the cell and translation for the assembly of proteins were described. The triplet genetic code underlying protein synthesis was revealed in 1966 by Robert Holley, Har Gobind Khorana and Marshall Nirenberg (Wagner, 1994). Finally, in the last quarter of the 20th century there were major developments in recombinant DNA technology, which resulted in a rapid advance in the knowledge of genes and their organisation within the genome. This progress led to the development of the human genome project in the 1980s.

The initial strategy of the project was to map the genome, using both linkage maps of polymorphic markers at around 2-5 cM intervals and high resolution physical maps of all 24 chromosomes by the construction of contig maps of overlapping clones (Olson, 2001 and McPherson *et al.*, 2001). The international human genome sequencing consortium then selected clones from this whole genome bacterial artificial chromosome (BAC) physical map to be sequenced and to produce the draft sequence of the whole genome. It was ensured that there was at least a seven-fold sequence coverage of each region (Lander et al., 2001).

The draft sequence of the human genome produced from pooled DNA from many anonymous individuals, was published by the international human genome sequencing consortium, a collaborative of twenty groups in United States, United Kingdom, Japan, France, Germany and China in February 2001 (Lander et al., 2001). In addition the private company Celera Genomics published a working draft of the

DNA sequence in the same month (Venter et al., 2001). At the time the haploid genome of 3.2 $x10^9$ base pairs (bp) (Alberts et al., 2002) was 25 times larger than any other genome sequenced. Repeat sequences that do not encode proteins make up at least 50 % of the whole genome (http://www.ornl.gov/hgmis/project/info.html). Indeed the size, complexity and large number of repeats within the human genome presented many challenges for sequence assembly (McPherson et al., 2001).

The draft sequence covered around 94 % of the genome and over 30 % was high quality finished sequence. It was estimated that the total number of genes is around 30-35,000, only about twice as many as a worm or fruit fly. However in humans there is believed to be more alternative splicing of transcripts, leading to an increase in the number of different protein products (Lander et al., 2001). It was fascinating to learn that the largest known human gene is dystrophin at 2.4 million bases and 178 exons (Alberts et al., 2002). It is hoped that the entire working draft will be finished to a high quality during 2003, coincidently the 50-year anniversary of the discovery of DNA (http://www.ornl.gov/hgmis/project/info.html).

Interestingly 99.9 % of all nucleotide bases are identical in all people (http://www.ornl.gov/hgmis/project/info.html). One class of sequence variation is Single Nucleotide Polymorphisms (SNPs) and they are especially valuable for understanding phenotypic variation between individuals, in particular for disease susceptibility and even for tracing human history. Sachidanandam *et al.* (2001) published a map of 1.42 million known SNPs at a density of 1 SNP per 1.91 kb and they describe how genetic diversity can be characterised not only in terms of individual polymorphisms, but also as specific combinations of alleles (haplotypes) at closely linked sites. The coverage of SNPs across the genome was found to be uneven, with SNP deserts and long regions of linkage disequilibrium representing a few haplotypes (Cardon and Bell, 2001). Indeed it is becoming more common to utilise haplotype blocks in genome-wide linkage disequilibrium (LD) association studies, rather than classical linkage mapping to locate genes/loci controlling complex traits (Cardon and Bell, 2001). In humans these markers are ideal in the

search for genetic loci controlling trait variation in an out-bred species where experimental crosses are not possible (Chakravarti, 2001).

The trait of most interest to humans and medicine is susceptibility to disease. The variation in genome sequences underlies differences in susceptibility to many types of diseases, age of onset, severity of the illness and the way our bodies respond to treatment. Prior to studies utilising genomics, variation in disease traits tended to be investigated for one gene at a time, with the gene typically chosen on physiological grounds (see section 1.3.2). However many diseases are complex and are controlled by several genes and therefore it is now possible to compare the patterns and frequencies of SNP genotypes in patients and controls and to determine which SNPs are associated with specific diseases (Collins et al., 1997). Ultimately this research may lead to the use of "personalised genetic medicine", a prospect which will dramatically change all aspects of the treatment of inherited diseases (Chakravarti, 2001).

The future of genomics in humans and other species lies in comparative and functional genomics studies. The transfer of information from humans to less well studied species and also the comparison of human sequence patterns to those of well studied model organisms, for example mice; are both very powerful approaches for the identification of genes across species and for determining their function. In addition gene-expression studies involving transcriptomics and proteomics allow the function of active genes to be understood and when, where and under what conditions the genes are expressed. This understanding will be particularly important in drug design studies.

1.1.2. Mice

An understanding of the genome of the laboratory mouse (*Mus musculus*) has many potential benefits. The mouse has been one of the most studied animals and has often been used as an experimental model for understanding the genetic control of many human traits, in particular disease susceptibility. Key advantages of this species are that controlled breeding populations can be created very rapidly and also

mutations can be created deliberately either by random mutagenesis or more importantly by directed engineering of the genome through transgenic, knockout or knockin techniques (Copeland et al., 2001). The effect of these mutations on the physiology and behaviour of the animals can then be studied (Coghlan and Cohen, 2002). The first transgenic mouse was created in 1982 by a team led by Richard Palmiter and Ralph Brinster. They fused elements of a gene that can be regulated by dietary zinc to a rat growth-hormone gene, and injected it into fertilized mouse embryos. The resulting transgenic mice, when fed with extra zinc, grew to be considerably larger than controls (Palmiter et al., 1982).

A high quality draft sequence of the mouse genome was published in 2002 (Waterston et al., 2002). 95 % of the total genome was sequenced and its overall size was estimated to be 2500 Mb (Coghlan and Cohen, 2002), that is around 700 Mb smaller than the human. The difference in size between the two genomes possibly reflects a higher deletion rate in the mouse lineage (Waterston et al., 2002). The sequence was assembled from pooled DNA from female mice of the C57BL/6J strain and had roughly a seven-fold coverage {Waterston, 2002 375 /id}. Therefore the Y chromosome still remains to be sequenced (Coghlan and Cohen, 2002). In addition around 80,000 SNPs have been revealed so far, by comparing the sequence with other mouse strains. These have revealed that genetic variation among strains occurs in large blocks, mostly reflecting the contributions of the two subspecies to laboratory strains, *Mus musculus domesticus* and *Mus musculus musculus* (Waterston et al., 2002).

By using comparative genomics this mouse genome sequence provides vital information for understanding the human genome, in much more detail than is possible with the human sequence alone. Indeed one such comparison revealed 1200 previously unidentified human genes. In addition many small segments of the mouse and human genomes, that do not contain protein-encoding genes, have been shown to be conserved. This preservation throughout evolution suggests that these segments play important roles and it is possible that they may be responsible for maintaining the structure of chromosomes or regulating gene expression (Coghlan and Cohen,

2002). Because the two genomes are distant in evolution, the divergence rate is high enough to be able to separate out functionally important elements by their degree of conservation (Waterston et al., 2002).

Even though mice and humans are separated by around 100 million years of evolution (Janke et al., 1994), the mouse genome has also been predicted to contain around 30,000 genes and only 300 genes (~1 %) are unique between these species. These 300 are mainly immunity, reproduction, detoxification and olfactory genes (Coghlan and Cohen, 2002).

It is certainly very interesting and useful to compare genome structure and sequences across species. Indeed researchers now plan to sequence the genome of a species that is the closest relative to the human, the chimpanzee (*Pan troglodytes*), in order to achieve a better understanding of our own genome and evolutionary history (Olson and Varki, 2003). The last common ancestor of chimpanzees and humans existed only around 5 million years ago and it is hoped that sequence comparisons will help us to understand what types of genetic change account for the emergence of a new species (Olson and Varki, 2003).

1.1.3. Livestock

Genome projects are in progress for all the major livestock species including pigs, chickens, cattle and sheep in both Europe and the USA. The traits of most interest in genomics studies in humans concern susceptibility to disease. However in livestock it is performance traits such as meat quality, reproductive efficiency and milk composition that have been studied in most detail. Although more recently the emphasis of research is moving towards more complex traits that traditionally breeding companies have had great difficulty selecting for, such as behaviour and disease resistance.

The majority of these economically important livestock traits display quantitative variation within populations or between lines or breeds. This variation is said to be quantitative because it cannot be classified into discrete classes in Mendelian

proportions; the trait displays a continuum between high and low performing animals (Haley and Andersson, 1997). This type of variation is usually controlled by several loci throughout the genome and the individual loci are known as quantitative trait loci (QTL).

There are two Mendelian laws of inheritance, which state that each allele has an equal chance of being passed onto the next generation ('equal segregation' law) and that an allele transmitted for one gene has no influence on those transmitted for other genes ('independent assortment' law) (Archibald and Haley, 1998). The fact that groups of genes are found together on chromosomes challenges Mendel's assumptions of the independence of genetic factors and genetic linkage clearly deviates from these laws. The closer the markers are together, the less likely they are to be separated by recombination at meiosis and the more likely they are to be genetically linked (Archibald and Haley, 1998).

Linkage mapping involves following the separation of marker alleles through the generations. This is done in order to establish whether or not alleles at a single locus co-segregate with alleles at other loci (Archibald and Haley, 1998). Where a QTL for a particular trait is closely linked to these genetic markers, the alleles at that marker will appear to be associated with differing levels of performance in the trait of interest.

A common strategy of mapping projects in livestock has been to construct reference genome-wide linkage maps of genetic markers throughout the genome, using experimental crosses (Rose, 1991). These crosses tend to be between breeds divergent in the production trait(s) of interest, for example Meishan crossed with Large White pigs and broilers crossed with layer strains of chickens. The frequency with which marker alleles separate across the generations, as a result of recombination, is calculated. This recombination distance (recombination fraction (r.)) is measured in centiMorgans. Where two markers are 1 cM apart, there is a one percent chance of recombination between them. The relationship between genetic and physical distances varies across the genome, but on average in mammalian

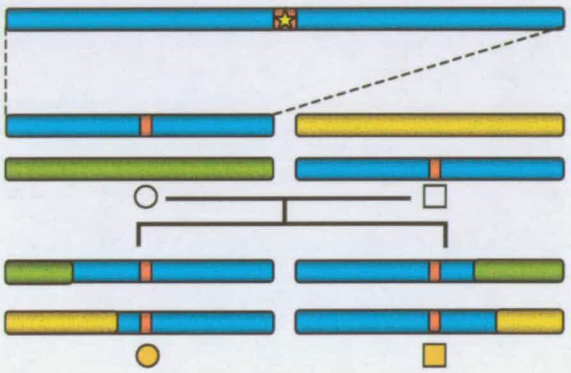
genomes 1 cM corresponds to 1 Mb. Although the chicken genome is about one third the size of the human genome in physical terms, it is of similar size in recombination terms. The average value across the eight macrochromosomes and the sex chromosomes Z and W is around 400 kb/cM. The overall physical size of the chicken genome, including the 30 cytologically indistinguishable microchromosomes is around 1200 Mb and the genetic size is around 3800 cM (Smith and Burt, 1998).

Polymorphic genetic markers are essential for the construction of these genome linkage maps and for the subsequent QTL mapping studies (Archibald et al., 1995). Over recent years there have been major advances in marker technology. Any identifiable segment of DNA in the genome, which shows variation between animals, can be used as a marker and there are now abundant, evenly distributed markers throughout the genomes of several livestock species, which are easily genotyped. The first markers used in genome studies involved the use of restriction enzymes to cut the DNA at specific sites, producing fragments of DNA of varying length between individuals, where there is variation in the sites recognised by the specific enzyme(s). These are known as restriction fragment length polymorphisms (RFLPs) and the 'marker' fragments (alleles) of DNA can then be used to distinguish different genotypes and to follow the pattern of inheritance across generations (Miller, 1997).

These RFLPs have now been mainly replaced by polymerase chain reaction (PCR) based markers, such as microsatellites. A microsatellite consists of a sequence of two to five base pairs long, repeated several times end to end, at specific sites throughout the genome. These highly polymorphic markers are abundant throughout the genome, for example there are an estimated 65,000 – 100,000 loci throughout the porcine genome (Archibald et al., 1995). Specific markers are chosen for experimental work on the basis of the degree of heterozygosity that is displayed in the parental generation. Informative markers tend to be selected where they have a level of heterozygosity of ≥ 0.5, where 1 represents a fully informative marker and 0 represents a non-informative marker (Haley and Archibald, 1992). Within a fully informative family, all parental alleles can be distinguished and all alleles can be traced through to the next generation (Lynch and Walsh, 1998).

Due to the longer generation intervals in livestock species, especially cattle, than for example in mice; the experimental crosses tend to consist of just a few generations from the founder animals. Therefore the amount of recombination is low and the regions of linkage disequilibrium around the QTL tend to be large i.e. in the order of several centiMorgans (cM). Figure 1-1 demonstrates the difference between the traditional linkage analysis carried out on structured experimental crosses and association analyses in natural populations for the identification of a causal mutation.

A. Linkage



B. Association analysis

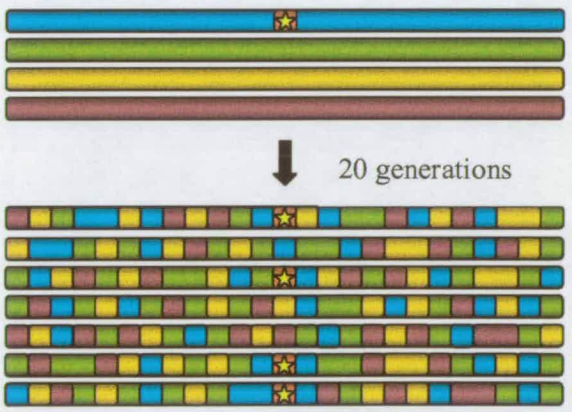


Figure 1-1 Schematic of the difference between a) linkage and b) association analyses. The region highlighted with a yellow star represents the functional mutation. Due to a smaller number of recombination events in linkage studies, the haplotype around the mutation will be much larger than found in association studies. In outbred random mating populations the mutation has been separated from the original haplotype over many generations. Adapted from (Cardon and Bell, 2001).

It can be seen that with association studies from natural breeding populations, the mutation tends to be separated from other alleles of the original haplotype over several generations of recombination. Therefore the region associated with the trait tends to be very small, even sometimes just the gene region itself containing the causal mutation (Cardon and Bell, 2001). However the mapping of QTL using linkage studies is a crucial start point in the identification of these causal mutations, in particular in those species where structured experimental crosses are possible. Especially as these QTL regions can now be compared to homologous regions in more studied species such as humans and mice, where many genes have already been identified and mapped.

As well as producing linkage maps, detailed physical maps of livestock species are also being developed, mainly from the bacterial artificial chromosome (BAC) libraries that are available (for example Anderson *et al.*, 2000 and (Warren et al., 2000). The aim of physical mapping is to locate genes or markers on specific regions of chromosomes. The lowest resolution physical map, a chromosomal map, is based on the banding patterns seen when stained chromosomes are observed under the light microscope (Chowdhary, 1998). In addition radiation hybrid panels are available for many species that allow high-resolution maps of the whole genome to be rapidly built. It is likely that in the future whole genomes of livestock will be sequenced from the BAC clones. For example a proposal has recently been put forward to sequence the pig genome (Rohrer et al., 2003). The pig has many similarities in structure and function with humans including their size, digestive physiology and type of diet (omnivores), propensity to obesity and they also display similar social behaviours (Tumbleson and Schook, 1996). This species has therefore often been used as a model to help understand human physiology, behaviour and onset of disease.

The porcine genome is also of a similar size and structure to that of the human (Archibald, 1994). It is therefore believed that not only will it be useful for studies of porcine genetics alone, it will serve as a reference non-primate, non-rodent, eutherian genome. This will allow a much better understanding of the genetic

control of human phenotypes through the knowledge acquired from controlled experimental studies undertaken in the pig, especially for disease traits such as diabetes and also traits associated with obesity (Rohrer et al., 2003). It is hoped that during 2003 the genome sequence of the chicken will be completed and the company Metamorphix Inc currently have 1x sequence coverage of the porcine and cattle genome and in addition sequencing of the canine genome is also now underway (Alan Archibald, personal communication).

1.1.3.1. Pig genome mapping projects

The pig (*Sus scrofa*) is ideal for genome mapping due to its well defined karyotype, large full-sib families, short generation interval (around one year) and availability of diverse genetic stocks (Archibald and Haley, 1990). The porcine genome consists of about three billion DNA base pairs contained within 18 pairs of autosomal chromosomes plus the sex (X and Y) chromosomes (Haley and Archibald, 1999). Figure 1-2 shows the porcine karyotype, which demonstrates the relative size of each of the chromosomes. The Giesma stain used to visualise the chromosomes reveals a G-banding pattern. The regions which are darkly stained generally correlate with DNA that is low in GC content and this indicates a lower density of genes (Alberts et al., 2002).

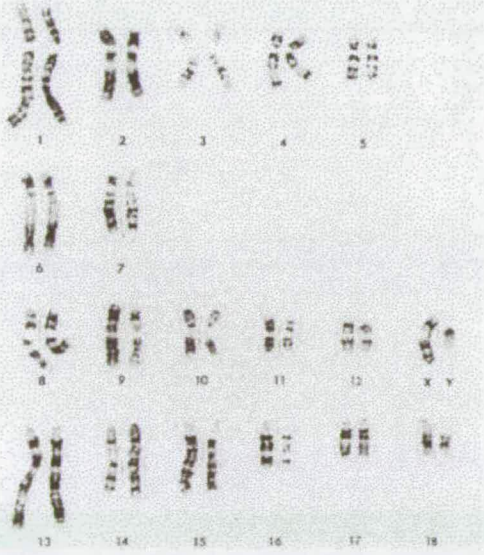


Figure 1-2 Photo of metaphase chromosomes stained to reveal the G-banding pattern of the porcine karyotype (courtesy of Roslin Institute). Each pair of homologous chromosomes is labelled.

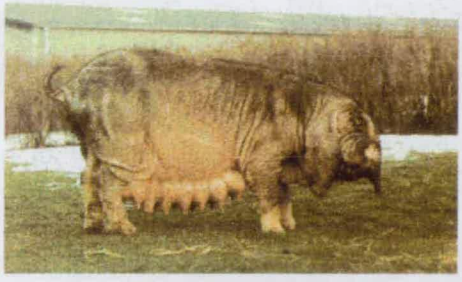
The physical size of the porcine genome has been estimated at around 2770 Mb (Schmitz et al., 1992), the equivalent genetic distance being around 30 Morgans (Archibald, 1994). This is a similar size to the 3200 Mb estimated for humans (Alberts et al., 2002) and it is believed that the pig also has around 30,000 genes (Archibald, 1994).

Characterisation of the pig genome was initiated in the early 1990s by a number of international consortia, as well as individual laboratories. For example the Pig Gene Mapping Project (PiGMaP) consisted of European laboratories, including the Roslin Institute in Edinburgh and the Nordic group bought together scientists in Uppsala, Copenhagen and Oslo. Major individual laboratories included scientists based at the Meat Animal Research Center in Nebraska.

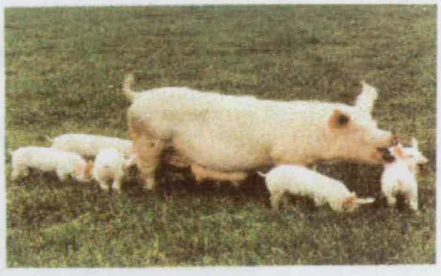
The initial aim of the PiGMaP group was to locate molecular genetic markers (recognisable sequences of DNA bases) that are roughly evenly spaced throughout the whole porcine genome. These markers were then used to construct genetic and physical maps of the genome. The genetic map produced of the male pig was approximately 16.5 Morgans, compared to the female map of around 21.5 Morgans (Archibald, 1995). As with all species studied, the recombination or genetic map of the heterogametic sex is shorter than that of the homogametic.

These genetic maps can be used to aid the understanding of the organisation and action of genes controlling valuable quantitative traits, to allow for the isolation or manipulation of these genes. The maps can also be used for marker-assisted selection or introgression of genes, allowing for the direct improvement of breeding stock and also facilitating the study of evolutionary relationships between species.

Three main types of breed were used to create the PiGMaP reference families. These are European commercial breeds (such as the Large White), the Chinese Meishan and European wild boar. Distinct genetic differences can clearly be seen between the three breeds as shown in Figure 1-3.



Chinese Meishan



Large White

Wild Boar

Figure 1-3 Chinese Meishan, European Wild Boar and Large White breeds of pig.

The Large White is a much more 'efficient' producer than the Meishan and is consequently used as a commercial breed worldwide. The European wild boar, having not undergone artificial selection, is very different from modern European breeds and from the Chinese Meishan. The genetic differences between these breeds is the key to locating genes, which control the variation observed in economically important traits, such as growth rate, meat quality and litter size (Archibald, 1990).

1.2. Marker assisted selection programs

Artificial selection of domesticated species has been practised for centuries, with the breeding of animals with the most desirable phenotypes (Lande and Thompson, 1990). It is relatively simple to improve livestock performance by altering the direct environment, however the resulting changes tend to be expensive and only apply in the short term. On the other hand genetic improvements are permanent and require virtually no maintenance. Once beneficial alleles at specific loci for traits of economic value have been shown to be associated with a positive improvement in those traits, it is possible to introduce marker-assisted selection (MAS) or marker-assisted introgression (MAI) within commercial breeding populations (Georges,

1999). Specific genes identified by positional cloning or within QTL regions can be used as markers to select between or within breeds. Marker-assisted selection describes where beneficial alleles are selected within the population and the frequency of that allele changed and marker-assisted introgression is where new alleles are introduced from another breed or population (Olliver, 1998). An example of marker-assisted introgression involved the introduction of the halothane resistance gene from the Large White breed of pig to the Pietran breed (Hanset et al., 1995).

By integrating molecular genetics approaches into traditional selective breeding methods, it should be possible to maximise improvement in the economic value of domesticated livestock populations (Lande and Thompson, 1990). The major benefit of MAS is to be able to increase selection intensity and to have the potential to produce an animal with favourable alleles at several loci controlling one or more traits of interest. However, large-scale genotyping of piglets or even embryos is a significant undertaking and costs are currently too high for widespread typing of multiple genetic markers (Georges, 1999).

The effect of MAS is potentially greatest for low heritable and sex-limited traits, where there has been limited success with traditional phenotypic strategies (Lande and Thompson, 1990). However it must be noted that for MAS to be successful, it is important that beneficial alleles for a given trait are shown to have no deleterious pleiotropic effects on other performance traits (Rothschild et al., 1997). Single genes of major effect that have been amenable to genetic engineering usually have been shown to have such deleterious effects. This explains why evolution in natural populations usually proceeds in a series of small steps (Fisher, 1958).

1.3. Approaches towards locating specific trait loci

1.3.1. QTL scan followed by positional cloning

The limitation with the identification of QTL regions alone is that the markers associated with a certain trait in one population raised under specific environmental conditions, may not be associated with the trait in another population with a different

genetic background and recombination history. It is also possible that the allele identified at the marker to be favourable in one population may in fact be unfavourable in another. This therefore limits the success of marker assisted selection programs with such markers. However once the causal mutation within the functional gene has been identified and verified, this mutation can be used for selection within most populations.

Therefore the natural follow-up to a QTL study is to attempt to locate the causal mutation by positional cloning and to carry out functional studies, for example to reveal differential expression patterns, in order to verify the causal nature of the variation. However locating causal mutations is a very difficult task, because the mutation within the QTL does not normally cause a major disruption in gene function, but instead the mutations cause subtle structural or regulatory changes (Coppieters et al., 1999).

The first report of positional cloning of the trait gene within a QTL in an out-bred mammal was the identification of a missense mutation in the bovine *DGAT1* gene with a major effect on milk yield and composition (Grisart et al., 2002). A QTL region of around 3 cM had previously been mapped to the telomeric end of bovine chromosome 14 (Farnir *et al.*, 2002; Heyen *et al.*, 1999; Looft *et al.*, 2001; Riquet *et al.*, 1999 and (Coppieters et al., 1998). A bovine BAC library (Warren et al., 2000) was then screened with microsatellite markers within the QTL and human cDNA clones that mapped to the orthologous chromosome region in the human. The ends of each BAC clone isolated were sequenced and sequence tagged sites (STSs) developed. These STSs were then mapped alongside the microsatellite and gene markers onto a bovine x hamster whole genome radiation hybrid panel (Womack et al., 1997) to confirm that they mapped within the QTL region and to identify their relative positions. A BAC contig spanning the QTL region was then constructed and aligned with the orthologous human HSA8q24.3 genomic "golden path" sequence (Lander et al., 2001) to highlight all the known genes in the region.

The gene *DGAT1* (encoding a protein with acylCoA:diaclglycerol acyltransferase activity) was predicted to lie within the QTL region and was considered a strong positional and physiological candidate gene. The QTL affected milk fat content and DGAT1 is known to catalyse the final step in triglyceride synthesis, with triclycerides being the major component of milk lipids. In addition when the gene was knocked out in mice, lactation was completely inhibited (Smith et al., 2000). From mouse and human cDNA sequences, Grisart *et al.* (2002) identified three bovine expressed sequence tags (ESTs) within the homologous gene and used these to identify the intron-exon boundaries. In a previous linkage study thirteen heterozygous sires for the QTL had been shown to transmit an allele, which markedly increased the milk fat content to half of their daughters, compared to inheritance of the alternative allele. Grisart *et al.* (2002) subsequently designed primers to amplify and sequence the coding region of exon 1, exon 2 and the whole region from exon 3 to exon 17 of the *DGAT1* gene from genomic DNA of the 13 sires. Four polymorphisms were revealed in these regions and a mutation at a double SNP site from bases *AA* to bases *GC* in exon 8, which causes an amino acid substitution from lysine to alanine and the allele encoding the amino acid alanine, was demonstrated to be the predicted causal mutation, affecting the function of the enzyme and resulting in an increased milk fat content.

All other studies to date, where a causal mutation has been identified for a trait of interest, the underlying genes were major genes that could be detected by Mendelian inheritance. Therefore the mutation was initially identified within a population and the gene located by various mapping strategies. The main difference with the study by Grisart *et al.* (2002) is that they identified a causal mutation within a QTL region.

One of the earlier examples of positional cloning of a causal mutation was the identification of a single point mutation in the porcine gene for the skeletal muscle ryanodine receptor (*RYR1*) associated with malignant hyperthermia (Fujii et al., 1991). Previously Davies *et al.* (1988) had located the halothane locus to porcine chromosome 6. Malignant hyperthermia is an inherited myopathy triggered by inhalational anaesthetics and this is why it was originally known as the halothane

gene. In pigs, it is also called porcine stress syndrome, as it is triggered by stress and results in pale, soft, exudative pork due to post-mortem manifestations of the disease in susceptible animals.

Kambadur *et al.* (1997) identified mutations in the myostatin (*GDF8*) gene related to double-muscling in Belgian Blue and Piedmontese cattle. This muscular hypertrophy (*mh*) phenotype results from almost a doubling in the number of muscle fibres and the autosomal recessive *mh* locus controlling this condition had been mapped by Charlier *et al.* (1995) to the centromeric end of bovine chromosome 2. The locus was then refined by Casas *et al.* (1998) to a 3-5 cM interval using a comparative mapping approach and the myostatin gene mapped within this interval (Smith et al., 1997). This gene is known to be expressed in adult skeletal muscle as well as during development, however Kambadur *et al.* (1997) demonstrated that there was no difference in the level of expression of the gene between normal and Belgium blue animals at various stages of development.

The key finding was that mice with a copy of the *GDF8* gene with exon 3 missing, the part of the protein which is highly conserved among the TGF- β super family of genes, displayed a phenotype similar to double muscling in cattle (McPherron et al., 1997). Kambadur *et al.* (1997) obtained a cDNA clone of bovine myostatin from a normal animal and compared it to mouse cDNA sequence and confirmed that the features common to the TGF- β super family were also conserved in cattle. A number of different mutations in the cattle *GDF8* gene have been identified. In the Belgium Blue breed an 11 bp deletion in the open reading frame was detected, which causes the loss of three amino acids and a frame shift resulting in a premature stop codon and a truncation of the protein, such that the conserved TGF- β region is not translated. In the Piedmontese cattle, a single base pair mutation results in a change from a cysteine to a tyrosine, which alters one of the conserved residues. It therefore appears that it is reduced activity of the resulting mutants of myostatin that results in a loss of control of muscle growth and that myostatin is a negative regulator of muscle growth in cattle and mice (Kambadur et al., 1997).

A mutation in the callipyge gene, which causes muscular hypertrophy, with associated leanness and improved feed efficiency, has recently been identified in sheep. Expression of the callipyge phenotype is the only known case in mammals of paternal polar overdominance gene action, i.e the only genotype that expresses the muscular hypertrophy is where the mutant callipyge allele is inherited from the sire and a normal allele from the dam (Freking et al., 2002). Cockett *et al.* (1994) had used comparative mapping to locate the locus to ovine chromosome 18. A single base mutation responsible for the callipyge phenotype was subsequently identified, but its mode of action remains unknown (Freking *et al.*, 2002). The detection of the mutation was from an inbred ram of callipyge phenotype postulated to have inherited chromosome segments identical-by-descent.

1.3.2. Physiological candidate gene approach using association studies

An alternative approach for identifying loci underlying traits of interest is to test known physiological candidate genes within association studies. There have been several recent examples of such studies of single candidate gene loci, in livestock species, for traits of economic importance. These include the association of a polymorphism within prolactin receptor gene (*PRLR*), which maps to porcine chromosome 16 and litter size in a Polish Landrace breed (Kmiec et al., 2001) and PIC lines of Landrace, Duroc, Landrace/Pietran and Meishan/Large White origin (Rothschild et al., 1998). In the study by Kmiec *et al.* (2001), there was seen to be an allele substitution effect at a SNP encoding an amino acid change, with a difference of 0.35 piglets between the two classes of homozygotes ($P < 0.1$). However the standard threshold of statistical significance is usually $P < 0.05$ and therefore this association could be questioned. In addition a one-way ANOVA was used and although the data were corrected for the year of birth of the litter, no additional influential factors on the phenotype were included in the analysis. Rothschild *et al.* (1998) also found a relatively small difference between homozygous sows of 0.25 piglets.

Interestingly Drogemuller *et al.* (2001) found a difference of more than one pig per litter between the two homozygous genotypes for a polymorphism within *PRLR* gene of a Duroc line. However the association analysis showed no significant association of the *PRLR* locus with litter size. In addition van der Steen *et al.* (1997) identified a significant allele substitution effect of a microsatellite marker within the gene, secreted phosphoprotein 1 (*SPP1),* and litter size in pigs. Van der Steen *et al.* (1997) describe an additive effect of between 0.41 and 0.95 piglets per allele copy. They suggested that the *SPP1* microsatellite marker must be linked to a QTL for litter size and that if *SPP1* is not the causal gene itself, then there could be a causal locus in LD with the microsatellite.

To date the locus found to display the most significant effect on litter size in pigs was a polymorphism within the oestrogen receptor gene (*ESR*) located on porcine chromosome 1. Sows of 50 % Meishan origin (PIC line) homozygous for the beneficial allele farrowed an extra 2.3 piglets in the first parity and an average of 1.5 piglets across subsequent parities (Rothschild et al., 1996). Short *et al.* (1997) confirmed the finding in a study of 4,262 sows with a total of 9,015 litter records. They found that the total number of piglets born was increased by 0.42 piglets per favourable allele of *ESR*. Gibson *et al.* (2002) tested the same polymorphism as described by Rothschild *et al.* (1996), with various sow productivity traits including ovulation rate and litter size, in an independent Meishan x Large White population. However in this population the effect of genotype at the bialleic polymorphism tested within *ESR* was found not to have a significant effect on any of the traits tested. In addition Rohrer *et al.*, 1999 and Drogemuller *et al.*, 2001 also failed to detect an effect of this specific polymorphism within *ESR* on litter size in their Meishan x Large White and Duroc x Large White populations respectively.

Due to the variation in association seen between populations, it is likely that the polymorphism identified within an intron of *ESR* by Rothschild *et al.* (1996) and Short *et al.* (1997) is in linkage disequilibrium with the causal mutation influencing litter size within the specific populations that they studied. As mentioned in section 1.3.1 it is vital to identify the causal mutation within the functional gene in order to

use this information for targeted selection programs within various pig breeds and populations. Keightley and Knott (1999) describe an analysis which tests whether QTL identified for similar traits across two independent studies are in fact the same QTL segregating in both populations.

Another example of a study in pigs was the association between leptin and back fat. Leptin mRNA levels are greater in adipose tissue from obese pigs than lean pigs (Robert et al., 1998) and therefore this gene was believed to be a strong candidate for the control of production traits such as back fat thickness and growth rate. Indeed two polymorphisms within the leptin (*LEP*) gene were found to be associated with feed intake ($P < 0.01$) and growth rate ($P < 0.01$) traits in the Landrace breed (Kennes et al., 2001).

In Holstein cattle an association was found between the growth hormone (*GH*) and growth hormone receptor (*GHR*) loci and milk yield traits (Vukasinovic *et al.*, 1999 and (Aggrey et al., 1999). Aggrey *et al.* (1999) tested three polymorphisms within *GHR* for association with the breeding values for milk yield and fat and protein content in 301 Holstein bulls using least squares methods. Bulls homozygous for one of the alternative alleles at one of the polymorphisms had a higher estimated breeding value (EBV) for milk fat content than both heterozygous and homozygous bulls for the alternative allele ($P < 0.01$). At the second polymorphism, homozygous bulls for one allele had a higher EBV for both milk fat and protein content ($P < 0.01$) than heterozygous bulls and the third polymorphism showed no association with milk-related traits. Vukasinovic *et al.* (1999) tested 553 Holstein bulls and they suggest that the bovine growth hormone locus is linked to a QTL that effects the EBV of milk protein percentage.

A study by Gavora *et al.* (1991) found an interesting association between *ev* genes at over 20 loci throughout the genome and reduced egg production rate, weight and quality in White Leghorn chickens. The genomes of virtually all chicken populations contain DNA sequences orthologous to avian leukosis virus (ALV) DNA inherited as

stable genes; these genes are known as endogenous viral (*ev*) genes (Sabour et al., 1992).

Recently there have also been many examples of allele association analyses of single loci within candidate genes and human disease traits. One such example was described in a study by Vasseur *et al.* (2002), who investigated the genetic risk of type 2 diabetes in French Caucasians. They found a strong association between a haplotype of SNPs 5' of the coding region and also at least one non-synonymous mutation in exon 3 of the *APM1* gene from DNA samples of individuals with type 2 diabetes. Within the regression analysis used to test for association they corrected for the effects of age, sex and body mass index. The gene *APM1* encodes for adiponectin, which modulates insulin sensitivity and glucose homeostasis. The level of this protein in plasma is decreased in both type 2 diabetes and obesity.

However due to over interpretation of results and poor study design the literature is also teeming with reports of association studies mainly in humans, that either cannot be replicated or for which corroboration by linkage has been impossible to find (Cardon and Bell, 2001). It is vital that robust statistical analyses are carried out, which account for all factors of variation on the trait, such as environment and genetic background and do not only investigate the gene locus in isolation.

Many association studies tend to investigate single loci that have been identified as putative candidate genes. In contrast SNP-based genome-wide association analyses are possible and in theory relatively straightforward. There are still however many technical and statistical obstacles. The technique needs to be reliable and efficient to type a large number of markers over the individuals of interest. The difficulties encountered with the analysis is that it must be able to detect small contributions to the phenotype from several genes or loci, whilst allowing for the inherent false positive error rate involved when testing such a large number of markers. It is also important to consider the importance of gene-gene or gene-environment interactions on the control of complex traits (Cheung and Spielman, 2002). The problem with quantitative characteristics is that where genes or loci are identified to have a

relatively small effect on that trait it can be difficult to ascertain their exact role within what can be a complex pathway of physiological control.

1.4. Female reproductive traits

The reproductive productivity of livestock depends on a complex of physiological pathways. The main factors of female fertility are the oestrus cycle, ovulation rate, conception rate, embryo survival, number of offspring at parturition and the weaning to oestrus interval (Merks et al., 2000). All of these traits are heritable, although in general they tend to display a low level of heritability. For example in pigs the heritability of litter size has been estimated to be between 0.05 and 0.15 (Alfonso et al., 1997; Hanenberg et al., 2001; Johnson et al., 1999c; Kerr and Cameron, 1996). The estimate for embryo survival rate is similar at 0.14 and ovulation rate is higher at 0.24 (Johnson et al., 1999c).

Genetic improvement of sow fertility was only really considered after the late 1980s, prior to which there was seen to be very little response as a result of selective breeding alone (Ollivier and Bolet, 1981). The application of Best Linear Unbiased Prediction (BLUP) and the use of family information within large populations resulted in improvements of 0.1-0.2 piglets per year during the 1990s (Merks et al., 2000). An understanding of the genetics underlying these quantitative traits in all livestock species should allow targeted selection and more significant improvement than has been possible to date.

1.4.1. Mapping of quantitative trait loci (QTL) and candidate genes

1.4.1.1. Sheep

The Booroola fecundity gene (*FecB*) has been known for a few years to map to sheep chromosome 6 and was believed to lie close to the *SPP1* gene (Montgomery et al., 1995b). *FecB* is associated with an increased ovulation rate of around 1.3-1.6 ova per copy and a subsequent improvement in litter size of around one extra lamb

(Montgomery et al., 1992). Recently Wilson *et al.* (2001) and Souza *et al.* (2001) identified the causal mutation in the ovary expressed bone morphogenetic protein 1B receptor *(BMPR1B)* gene in Booroola sheep.

BMPR1B is the second gene discovered in the TGF-β pathway to affect ovarian function and fertility in sheep, the first being the Inverdale gene on sheep chromosome X, which displays a mutation in the coding region of the *BMP15* (*GDF9b*) gene (Galloway et al., 2000; Otsuka et al., 2000). A copy of the mutation results in an increased ovulation rate and twin and triplet births in heterozygotes, however homozygotes with two copies of the mutation display ovarian failure and infertility (Galloway et al., 2000). Indeed several members of the TGF-β super family, such as the inhibins and activins, Müllerian inhibiting substance, BMP8a, BMP8a and GDF9 have been implicated as important regulatory factors in mammalian reproduction (Dube et al., 1998).

Mandiki *et al.* (2000) identified a similar gene to the Booroola fecundity gene in the prolific Cambridge breed of sheep, where ovulation was increased in carriers. It has been named the Cambridge fecundity gene and it would be interesting to know whether it is due to the same causal mutation as the one found in *BMPR1B*.

Another gene for fecundity was proposed in Icelandic sheep, where all multiple births on a farm were traced back to a single ewe named "Thoka" (Jonmundsson and Adalsteinsson, 1985). This putative major gene (Thoka) was introduced into Cheviot sheep in the UK resulting in a mean litter size of 1.85 lambs per litter, as compared to 1.26 lambs reported for a normal Cheviot flock (Walling *et al.*, 2002 and (Russel et al., 1997). Walling *et al.* (2002) used segregation analyses to confirm the presence of a major autosomal gene for litter size in Thoka Cheviot sheep, with an additive effect of an increase in litter size of 0.70 lambs per copy of the gene.

The gonadotrophin-releasing hormone receptor gene (*GNRHR*) was mapped by Montgomery *et al.* (1995a) to sheep chromosome 6. GnRH binds to GNRHR on cells of the pituitary gland to stimulate the synthesis and release of follicle-

stimulating hormone (FSH) and luteinising hormone (LH), both key factors in the control of folliculogenesis (Montgomery et al., 1992). At one time *GNRHR* was believed to be a candidate gene for the Booroola mutation. However *GNRHR* was shown by Montgomery *et al.* (1995a) to map outside the region of the Booroola locus.

1.4.1.2. Cattle

Reproductive performance is particularly important to the cattle industry and Kappes *et al.* (2000) state that reproductive traits are twice as economically important as production traits. The desire is to select bulls that will sire daughters that produce twins rather than single calves. However there are a number of problems associated with twinning that need to be considered. These include a greater incidence of calf mortality, dystocia, calf abandonment and an increased occurrence of freemartin heifers. This is where the chorionic blood vessels fuse between two fetuses of each sex and as a consequence müllerian inhibiting substance is transferred from the male to the female and this suppresses the differentiation of the female reproductive tract and consequently the females are sterile (Kappes et al., 2000). It is likely that this issue is not so important in the beef industry as it might be in the dairy industry.

Twinning rate is basically a product of ovulation rate, conception rate and embryo survival. Blattman *et al.* (1996) carried out a genome-wide search for QTL for ovulation rate in cattle selected for increased twinning rate and found putative regions on chromosomes 7 and 23, with strongest statistical support for the QTL on BTA7. In an extension of this study, Kirkpatrick *et al.* (2000) mapped additional markers to chromosome 7 in an attempt to refine the QTL region and to search for multiple QTL on this chromosome. Using a genome wide scan, they identified regions on chromosomes 5, 7, and 19, that appeared to influence the estimated breeding value (EBV) for ovulation rate. They found no evidence for multiple QTL on chromosomes 5 and 19, however there did appear to be two QTL on chromosome 7. From a search of comparative human gene maps, Kirkpatrick *et al.* (2000) suggest that there are no obvious candidate genes controlling ovulation rate in the homologous regions to the QTL mapped to bovine chromosome 5. The gene for the anti-müllarian hormone was predicted to map to a QTL region on bovine

chromosome 7. The QTL on chromosome 19 maps above the gene for bovine growth hormone mapped by Kappes *et al.* (1997) and this gene may control bovine ovulation rate (Kirkpatrick et al., 2000).

Both Kappes *et al.* (2000) and Lien *et al.* (2000) report genome wide scans for traits affecting twinning rate in cattle. Kappes *et al.* (2000) studied a population that was selected for increased twinning rate and searched for loci or genes controlling ovulation rate. They typed 273 markers across all 29 autosomes and the X-chromosome and found the strongest evidence for a QTL that affects ovulation rate, significant at the genome-wide level, at around 40 cM from the centromeric end of chromosome 5.

Lien *et al.* (2000) measured twinning rate in a population of Norwegian cattle and used a genetic map of all 29 autosomes. They found potential QTL for twinning rate on chromosomes 5, 7, 12 and 23. As with Kappes *et al.* (2000), the QTL of highest significance was on chromosome 5 and Lien *et al.* (2000) also carried out a comparative analysis with the human genome to attempt to identify candidate genes in the QTL region. They predict that *IGF1* (insulin-like growth factor) maps within the QTL region and describe it as a strong physiological candidate, due to its role in stimulating progesterone production and mitosis of bovine ovarian granulosa cells. They subsequently searched for mutations within the gene in families segregating for the QTL and only found mutations in non-coding regions of the gene.

1.4.1.3. Mice

Kirkpatrick *et al.* (1998) carried out a QTL mapping study in both F2 intercross and backcross mouse populations from lines that differed markedly in prolificacy and mature weight. They found strong evidence for QTL controlling increased litter size in the F2 intercross population on chromosomes 2, 4 and 9.

Pomp *et al.* (1995) also investigated an F2 intercross population and found evidence for QTL with additive effects on chromosomes 3, 8, 11, 13, 18 and 19 for ovulation rate. They also found QTL for embryo survival on chromosomes 2, 9, 10, 11 and 19 and for litter size on chromosomes 2, 10 and 19. The QTL of largest effect was seen

on chromosome 2, where individuals inheriting both alleles from the selected line displayed an increase in embryo survival of 13.6 % and an extra two live pups.

Spearow *et al.* (1995) investigated QTL controlling genetic differences in hormone-induced ovulation rate in contrasting strains of mice. They mapped suggestive QTL on chromosomes 2, 4, 6 and 9. They recently mapped these QTL for ovulation rate at a higher resolution, confirming QTL in the regions of central chromosome 2, proximal chromosome 2, central chromosome 6 and chromosome X. All were highly significant, displaying additive effects of 2.4 to 6.8 ova per allele (Spearow et al., 2000).

1.4.1.4. Pigs

There have been several QTL scans for female porcine reproductive performance traits, many of which have highlighted effects on chromosome 8. Rathje *et al.*, 1997 conducted a genome wide scan over a ten generation index line of Large White x Landrace pigs, selected for increased ovulation rate and prenatal survival crossed with a ten generation random selected line. A significant QTL for ovulation rate was proposed at the telomeric end of the long arm of chromosome 8 (SSC8). This was close to the marker SW790 and displayed an additive effect of 3.07 ovulations from the index line. They found no significant QTL for litter size or the number of mummified pigs.

In a subsequent study on the same population, Cassady *et al.* (2001) increased the number of markers and analysed more animals and found evidence for a QTL towards the telomere of the q arm of SSC8 for age at puberty. However they could not confirm the QTL for ovulation rate on SSC8 found in the previous study, instead they identified a QTL for this trait on SSC9. They also found evidence for QTL for number of stillborn piglets on SSC5 and SSC15 and for teat number on SSC8 and SSC11.

Wilkie *et al.* (1999) also carried out a genome wide scan, but this time in a Meishan x Yorkshire pedigree population. A significant QTL for ovulation rate was detected around the centromere of SSC8, close to the marker SW444. The QTL displayed an

additive effect, which contrary to expectation displayed a negative contribution from the Meishan. They also found significant QTL for number of stillborn piglets on chromosome 4 and gestation length on chromosome 9.

Wilkie *et al.* (1999) also proposed suggestive loci for ovulation rate on chromosomes 15 and 7, gestation length on 15 and 1, uterine length on 7 and 5 and litter size on 6. In a follow up study by Braunschweig *et al.* (2001) more markers were scored across the same Meishan x Yorkshire population and the SSC8 centromeric QTL for ovulation rate was confirmed with increased confidence. However it is difficult to understand how the statistical support for a QTL can be improved by typing additional markers over the same individuals. The most efficient way to improve statistical power is to increase the number of animals in the study.

Milan *et al.* (1998) looked at 16 chromosomes and found some evidence for QTL affecting reproduction on porcine chromosomes 7 and 8. The traits included ovulation rate, number of embryos, weight and length of uterine horn and they proposed the associations on chromosome 8 with ovulation rate to be in the region of the telomere of the q arm. The estimated allele substitution effect ranged from 1 to more than 2 corpora lutea, with the most favourable alleles coming from the Meishan breed.

A more recent study by Rohrer *et al.* (1999) involved a multi-generation crossbred Meishan-Large White composite resource population. Again a genome wide scan was carried out and QTL for ovulation rate were detected on SSC3, 8, 9, 10 and 15. However in contrast to the other studies, the QTL on SSC8 was placed at the telomeric end of the short arm. As with the previous study by Wilkie *et al.* (1999), the alleles from the Large White were superior for all the QTL except the one on chromosome 15. Suggestive QTL for uterine capacity on chromosome 8 and ovulation rate and uterine capacity on chromosome X were also proposed.

Rohrer (2000) carried out a QTL scan for traits of birth characteristics and found significant evidence for QTL for teat number on porcine chromosome 10 and

suggestive evidence for QTL on chromosomes 1 and 3. The mode of gene action was additive for all three regions, with the Meishan alleles for the QTL on SSC3 and 10 increasing the number of teats and the Meishan alleles for the QTL on SSC1 reducing teat number. In addition Hirooka *et al.* (2001) found strong evidence for teat number QTL in a Meishan x Dutch cross on chromosomes 10 and 12, with the beneficial alleles from the Meishan breed and a QTL on chromosome 2, with a negative effect of the Meishan allele.

Jiang *et al.* (2002b) mapped QTL for prolificacy on SSC8 using a candidate gene approach. They claim that an allele predominant in the Meishan breed, within the fibrinogen G gamma polypeptide (*FGG*) gene, was associated with an increase in number of piglets born and number born alive for sows having their second parity. They also describe an association with a Meishan allele at the amphiregulin (*AREG*) gene and a decrease in the number of piglets born in the first parity and that the estrogen sulfotransferase (*STE*) gene showed overdominance effects on gestation length and number of teats. Previously Jiang *et al.* (2001) found an association with a single base substitution in the 5' UTR of the *GNRHR* gene and number of corpora lutea at the first parity of the Roslin Meishan x Large White population.

So far only one major gene for litter size has been identified. As mentioned in section 1.3.2, Rothschild *et al.* (1996), crossed Meishan and Large white breeds and demonstrated that an allele at the estrogen receptor (*ESR*) locus on pig chromosome 1, was associated with a significant increase in litter size. This finding was confirmed by Short *et al.* (1997). However Drogemuller *et al.* (2001), Rohrer *et al.* (1999) and Gibson *et al.* (2002) failed to detect an effect of this specific polymorphism within *ESR* on litter size in their Meishan x Large White and Duroc x Large White populations.

Janss *et al.* (1997) carried out segregation analysis between the same breeds, identifying a major gene affecting litter size at the first parity, which they believe is unlikely to be the *ESR* gene. This unknown gene was demonstrated to be dominant

and resulted in a difference of five or six piglets between homozygous recessive and homozygous dominant individuals.

1.4.2. Differences in physiology between pig breeds

The European commercial Large White or Yorkshire and the Chinese Meishan are genetically and phenotypically divergent. The Meishan pig is a member of the Taihu group of breeds from China and is one of the most prolific pig breeds known. This breed has litter sizes of three or more piglets greater than those of European breeds do (Haley et al., 1995). Therefore they are highly likely to carry alleles with the potential to enhance efficiency of production within commercial breeds such as the Large White (Haley and Lee, 1993). In order to attempt to understand the genetic basis of reproductive performance it is necessary to determine how the Meishan breed regulates its prolificacy.

Litter size is a product of ovulation rate (number of corpora lutea measured on both ovaries), prenatal survival and uterine capacity (Webel and Dziuk, 1974). Previously there has been uncertainty over the relative contributions of ovulation rate and prenatal survival to the subsequent litter size. Studies with Meishans have yielded conflicting data for ovulation rate (Cheng, 1983 and (Bidanel et al., 1989) but several have concluded that prenatal survival is an important factor influencing litter size (Haley and Lee, 1993; Bidanel *et al.*, 1989 and (Ashworth et al., 1997). Meishan gilts and sows have significantly higher ovulation rates (in excess of five ova) than Large White females of a similar age. When the breeds were compared after a fixed number of oestrus cycles, ovulation rates were found to be similar due to Meishan females being younger and smaller (Haley and Lee, 1993).

Interestingly the F1 cross between these 2 breeds displayed heterosis in prenatal survival and litter size, but not ovulation rate. That is, the crossbred (F1) animals showed higher levels of performance that either of the two pure breeds (Haley and Lee, 1993) (Table 1-1).

Table 1-1 Ovulation rate, number of live births and embryo survival (± SEM) in unilaterally ovariectomised-hysterectomised Meishan, Large White and F1 crossbred pigs (from (Haley and Lee, 1993). Note these values are for just one uterine horn.

Characteristic	Meishan	Genotype of female Large White	F1 crossbreed
Number of records	26	32	28
Ovulation rate (OR)	19.4 ±0.7[a]	15.0 ±0.7[b]	18.6 ±0.7[a]
Number of live births	7.0 ±0.6[a,b]	6.0 ±0.6[a]	8.2 ±0.6[b]
Prenatal survival (%)	37.6 ±4.0[a]	41.2 ±3.8[a]	47.6 ±4.4[a]
Prenatal survival (adjusted for OR)	42.0 ±3.8[a,b]	36.8 ±3.6[a]	49.2 ±3.9[b]

A restricted maximum likelihood analysis of first and second parity data was performed. Within rows, means with different superscripts are significantly different (P <0.05).

The Meishan has also been shown to display an increased uterine capacity, achieving this by a greater level of organisation in the uterus, that is more even spacing between attachment sites than the Large White (Lee et al., 1995). Uterine capacity is defined as the number of piglets, which a sow can maintain to term. It can be difficult to measure, especially in breeds where their ovulation rate is not great enough to challenge the capacity of the uterus. If the extra embryos present in Meishan sows are to result in extra piglets at birth, there must be a higher level of fetal survival.

Around 75 % of total porcine embryo loss occurs during the peri-implantation period (days 12-18 of gestation) (Ford, 1997). The Meishan breed has a marked reduction in conceptus loss during this period compared to the average 30 % loss observed in US and European breeds such as the Large White (Ford, 1997). Haley and Lee (1993) describe UK studies where the advantage to the Meishan breed at days 20-50 of gestation was seen to be a 19 % increase in survival levels compared to the Large White breed. This difference in embryo survival levels of the two breeds can be attributed to the smaller size of the Meishan embryo and more even spacing between attachment sites, as well as to the increased levels of placental efficiency of the Meishan sow compared to the US/European breeds (Biensen *et al.*, 1999; Ford, 1997; Vonnahme *et al.*, 2002; Lee *et al.*, 1995 and (Wilson et al., 1999).

From days 5-12 of gestation the Meishan embryos have a smaller number of trophectoderm (TE) cells than Yorkshire embryos, due to a lower mitotic rate of these cells (Ford, 1997). These TE cells produce an enzyme P450 17 α-hydroxylase/17-20 lysase, which is associated with oestradiol production by the conceptuses around day 12 of gestation. Therefore the Meishan embryo secretes less oestradiol into the uterine luminal fluid than the Yorkshire breed (Ford, 1997). The progesterone-primed endometrium of the uterus is sensitive to the conceptus oestradiol. This time of trophoblast expansion and attachment is temporarily associated with a down regulation of progesterone receptors and an up regulation of oestrogen receptors in the uterine epithelium (Johnson *et al.*, 2000; Lessey *et al.*, 1996 and (Geisert, 1997). In response to oestradiol, the endometrium alters its stromal and epithelial development as well as its secretory activity.

During the late peri-implantation period the blastocysts are dependent on the histotroph secretions from the endometrium for growth, development and elongation of the trophectoderm cells (Geisert and Yelich, 1997). As a consequence of the decreased secretory activity, the elongation of the conceptuses is less in the Meishan breed than in the Yorkshire breed (Wilson and Ford, 2000). Due to the more rapid growth of the Yorkshire blastocysts, the more advanced ones in the litter alter the uterine environment to the detriment of their less well developed litter mates (Wilson et al., 1999). More of the Meishan embryos survive the critical period of loss from days 12-18 because they develop more slowly and there is less competition between the embryos. In addition the rate of expansion of the trophoblast establishes the boundaries for placental attachment and initial allocation of uterine space available to each conceptus. Therefore a limiting factor in the Yorkshire breed is the increased amount of placental surface area required per fetus to obtain nutrients and oxygen for growth and survival to term (Geisert and Yelich, 1997 and (Wilson and Ford, 2000).

Another important strategy of the Meishan, which allows increased prenatal survival, is its high placenta efficiency (Wilson and Ford, 2000). In the later stages of gestation, the fetuses grow much more rapidly. In order to supply the increase in requirement of nutrients and oxygen the Yorkshire strategy is to increase the size of

each placenta and therefore increase the surface area of attachment of the fetus in the uterus (Vonnahme et al., 2002). In these European breeds there is a total loss of around 15-20 % of the fetuses around days 30-50 of gestation because the uterine space limits the increase in placenta size (Vonnahme et al., 2002). However, the Meishan strategy is to dramatically increase the vascular density to each attachment site and therefore maintain the small placenta sizes (Vonnahme et al., 2002). The Meishan placental weight and surface area does not increase from day 70 to term (day 114). Throughout gestation and at farrowing the Meishan conceptuses and placentae are smaller and lighter than US/European pig breeds (Biensen *et al.*, 1998; Lee and Haley, 1995 and (Wilson et al., 1998). In addition there has been observed to be a reduced variation in birth weight of Meishan litters compared to the Large White breed (Lee and Haley, 1995).

Experiments have been carried out where Meishan and Yorkshire fetuses have been cross-gestated in sows of both breeds in order to determine the effects of uterine environment versus fetal genotype on the differences in prenatal survival between these two pig breeds. It was shown that the Meishan uterus exhibits an inhibitory effect on the growth of the conceptus during the peri-implantation period (Biensen et al., 1999). Meishan and Yorkshire embryos gestated in a Meishan uterus to day 12 of gestation were smaller than embryos of both breeds gestated in a Yorkshire uterus (Biensen et al., 1999). In contrast, during late gestation the placenta size, vascularity and consequent placental efficiency appears to be determined by conceptus genotype. When Meishan and Yorkshire embryos were co-transferred to a Yorkshire uterus, the placenta size of the Meishan conceptuses were significantly lighter and much more vascular than the Yorkshire littermates (Wilson et al., 1998).

At day 90 of gestation, Meishan and Yorkshire fetuses in a Meishan uterus were a similar size and Meishan embryos were significantly lighter than Yorkshire when gestated in a Yorkshire uterus (Figure 1-4). Interestingly though by term, the growth rate of the Meishan embryos had accelerated such that the birth weights of the Yorkshire and Meishan littermates were similar (Biensen et al., 1998).

Figure 1-4 Litter of Meishan and Yorkshire fetuses co gestated through to day 90 in either a Meishan or Yorkshire uterus. The fetuses in the Meishan uterus are all of a similar size and there is a clear difference in size between breeds within the Yorkshire uterus. (Photos kindly supplied by Prof. Stephen Ford, University of Wyoming. Figure published in (Wilson et al., 1998)).

Galvin *et al.* (1993) also investigated the early post attachment period (20-22 days after mating) in Meishan and Large White gilts to determine whether any general breed differences were apparent at this time. They showed that Meishan sows were around 30 kg lighter, with 7.1 more ova, 8.7 more uterine horn attachment sites and 8.2 more embryos than Large White sows. The F1 sows were approximately intermediate in these measurements between the two breeds. It was also reported that the Meishan had slightly heavier (not significant) uteri than the Large White and the F1 sows had significantly heavier uteri than either pure breed. The uteri of the Meishan were the shortest and the F1 sows the longest. The embryonic and placental weights were least in the Meishan and greatest in the F1, showing significant evidence of heterosis.

Lee *et al.* (1995) used unilateral hysterectomy-ovariectomy methods as a more powerful tool to study the variation in uterine capacity. This is where one uterine horn with its ovary is removed at the bifurcation and as a result of hypertrophy of the remaining ovary, more than half of the ova crowd into a single uterine horn. In addition, the ovulation rate and number of embryos was in excess of the number born. Therefore it could be assumed that the uterine capacity of the horn was challenged. It was also noted that unlike intact females, there was no significant association between litter size and ovulation rate.

It was shown that in terms of total litter weight, the Meishan uterus had a lower capacity of around 3.5 kg than the Large White and the uterus of the crossbred could accommodate 3 kg more than the mid breed mean. Therefore the difference in uterine capacity between the two breeds does not appear to be a function of the physical size of the uterus (Lee et al., 1995). Embryos in Meishan sows had significantly less space to develop and as a consequence were lighter at the same comparative age as the Large White embryos (Haley and Lee, 1993).

In summary, the Meishan breed has reduced within litter variation in for example embryo length and weight and the distance between the embryo sites on the horn compared to the Large White. The Meishan has a greater capacity to maintain the fetuses to term and does so, not by the use of a larger uterus, but by a better organisation of embryo spacing and placenta efficiency. This in turn leads to more even fetal development and reduced competition between neighbouring fetuses for maternal resources (Lee et al., 1995).

Crossbred sows have physically longer uterine horns than Meishans and can therefore maintain a larger number of fetuses and in a more organised fashion than the Large White. This could explain the observed heterosis in litter size and prenatal survival. It is likely that several genes of important effect contribute to the different components of the reproductive performance of the Meishan pigs. In particular the genes, which control ovulation rate and fetal survival, subsequently controlling litter size. As seen in section 1.4.1.4 many studies have been undertaken to determine the

genetic component of the observed differences in reproductive performance between various breeds of pig.

1.5. Known physiological candidate genes on SSC8 controlling female reproductive traits

1.5.1. Gonadotrophin-releasing hormone receptor

Gonadotrophin releasing hormone (*GnRH*) is released from the hypothalamus and regulates the secretion of the gonadotrophins follicle stimulating hormone (FSH) and luteinising hormone (LH) from the anterior pituitary. These hormones are secreted in the plasma to the ovary where they stimulate antral growth and folliculogenesis in preparation for ovulation (Johnson and Everitt, 1995). *GNRHR* was mapped near the centromere of porcine chromosome 8 by Rohrer (1999). As mentioned previously Wilkie *et al.* (1999) and Braunschweig *et al.* (2001) mapped a QTL for ovulation rate to the centromere of SSC8. Also Jiang *et al.* (2001) revealed an association between a single base substitution in the 5' UTR of the *GNRHR* gene and number of corpora lutea at the first parity of the Roslin Meishan x Large White population.

1.5.2. SPARC-like 1

SPARCL1 (SPARC-like protein 1 or high endothelial venule protein) is an extracellular matrix glycoprotein, with a function in cell adhesion and proliferation and is believed to have anti-adhesive properties similar to those of SPARC. It is expressed in several tissues including ovary and lower levels in the placenta (Girard and Springer, 1995). Bertani *et al.* (2002) used differential display PCR to investigate differences in gene expression in the anterior pituitary of a group of control pigs and the Nebraska line selected for increased ovulation rate and embryo survival over several generations. They found that *SPARCL1* was differentially expressed between the two groups of pigs. They also mapped the gene within the QTL for age at puberty around the telomere of the q arm of SSC8 identified by (Cassady et al., 2001).

1.5.3. Secreted phosphoprotein 1

Secreted phosphoprotein 1 (*SPP1*), originally known as osteopontin (*OPN*) is principally produced by osteoclasts and osteoblasts and is found in abundance in the bone matrix (Denhardt and Noda, 1998). These bone cells are responsible for remodelling of mineralised tissues and it is believed that *SPP1* plays a role in cell attachment and controls bone cell functions such as resorption (Yoshitake et al., 1999). The main receptors for SPP1 are $\alpha_v\beta_3$ integrin heterodimers and are found on tissue cell surfaces, the SPP1 protein binds via an RGD (Arginine-Glycine-Aspartic acid) site (Johnson et al., 1999a). Wrana *et al.* (1989) and Hijiya *et al.* (1994) clearly demonstrated the position of the RGD binding site in exon 6 of the *SPP1* coding sequence in pigs and humans.

Although *SPP1* was originally discovered for its role in bone biology, it has also been shown to be expressed in a wide variety of cells including epithelial cells, predentin, kidney, thyroid, breast, testes, leucocytes, smooth muscle cells and high concentrations of the protein have also been detected in plasma of patients with disseminated carcinomas (Senger et al., 1989; Waterhouse et al., 1992). Of particular interest is the expression of *SPP1* by secretory phase endometrial cells, the placenta, invading trophoblasts and decidual metrial glands during blastocyst invasion and placentation (Johnson et al., 1999a), indicating that this gene has an important role to play in embryo implantation.

Factors shown to be responsible for increases in *SPP1* transcription include interleukin-1α and -1β, transforming growth factor (TGF), interferon gamma (IFN γ), oestrogen and progesterone (Kreiss and Vale, 1993). The release of progesterone and interferon-tau (the maternal recognition of pregnancy signal) from the trophoblast, during the peri-implantation period, have been shown to result in increased expression of *SPP1* mRNA from the luminal epithelium of the uterine endometrium in humans (Omigbodun et al., 1997), mice (Nomura et al., 1988) and pigs (Garlow et al., 2002), from the glandular epithelium of the uterus and decidualising stroma of baboons (Fazleabas et al., 1997) and from the glandular epithelium of ewes (Johnson *et al.*, 2000 and (Johnson et al., 1999b).

SPP1 codes for an acidic 70 kDa glycoprotein. This protein undergoes extensive post-translational modification, in particular heavy phosphorylation of serine and threonine residues (Safran et al., 1998). After treatment with proteases, the 70 kDa protein is cleaved to give 24 kDa and 45 kDa fragments. It is the 45 kDa fragments which contain the RGD binding site responsible for cell-cell attachments with integrin receptors (Johnson et al., 1999b). SPP1 binds primarily to αvβ3, αvβ5 and α4β1 integrin heterodimer receptors at the cell surface via its RGD sequence (Bowen et al., 1997; Johnson et al., 1999b; Johnson et al., 1999a; Johnson et al., 2001; Waterhouse et al., 1992). This biologically active 45 kDa form, is released as part of the histotroph secretion from the endometrium and has a high affinity for αvβ3 integrin (vitronectin) receptors expressed by the trophectoderm (the outer layer of the blastocyst) and the uterus during the implantation window. The 45 kDa protein is susceptible to degradation and it is thought that progesterone also regulates the production of tissue inhibitors of metalloproteinases, that prevent degradation of the active form of the protein and allow interaction of SPP1 with the integrin receptors (Johnson et al., 2000).

The binding of SPP1 to the vitronectin receptors on the trophectoderm and uterus stimulates changes in morphology of the trophectoderm and extra-embryonic endoderm that result in the cytoskeletal reorganisation and elongation of the conceptus. It also induces adhesion and cell signalling between the luminal epithelium and trophectoderm essential for attachment, superficial implantation and placentation (Johnson et al., 1999b; Johnson et al., 2000).

SPP1 plays a key role in conceptus implantation and maintenance of pregnancy in several known mammalian species. In sheep, days 11-17 of pregnancy incorporate the initiation of adherence and attachment phases of early implantation (Johnson et al., 1999a). There is an increase in SPP1 mRNA expression on day 13 and an increase in the 70 kDa and 45 kDa proteins in uterine flushings from pregnant ewes at day 15 compared to normal cyclic ewes (Johnson et al., 1999b).

The trophectoderm of day 19 conceptuses and luminal epithelial cells of the endometrium within the uterus are believed to express the integrin receptors, which bind the secreted SPP1. An increase in $\alpha_V\beta_3$ integrin expression in human endometrial epithelium during the peri-implantation period has also been demonstrated and β_3 expression shown to be reduced in infertile women (Lessey et al., 1996). Baboons also show co-expression of SPP1 and $\alpha_V\beta_3$ receptors in invading trophoblast and endometrial cells (Fazleabas et al., 1997). Additionally, mice injected with RGD peptides into the uterine lumen showed a decrease in implantation rate, due to an interference with these peptides and the $\alpha_V\beta_3$ receptors (Aplin, 1997). In mice SPP1-null mutants remain fertile (Liaw et al., 1998) whereas null mutations of β_1 or β_5 integrin genes leads to perinatal lethality (Hynes, 1996). Therefore in mice, it appears that SPP1 can be replaced by another ligand, binding to the same receptor and is therefore not essential for implantation. However *SPP1* obviously does play an important role, in several mammalian species, in implantation.

In pigs, an increased expression of *SPP1* mRNA has been shown in the uterine luminal epithelium, in regions of close proximity of conceptus tissue, after day 15 of gestation (peri-implantation period) and has been shown to result in integrin activation and the accumulation of the cytoskeletal molecules required to form the "focal adhesions" for adhesion and signalling between the conceptus and the uterus (Garlow et al., 2002). Expression of *SPP1* mRNA was also shown to increase significantly in the uterine glandular epithelium from day 35 to day 85 of gestation. Interestingly porcine conceptuses were not shown to express *SPP1* mRNA, however the protein was found at the conceptus trophectoderm as well as the uterine luminal epithelium (Garlow et al., 2002).

Omigbodun *et al.* (1997) proposed that in humans a paracrine regulatory feedback loop occurs within the villus of the trophectoderm (Figure 1-5). Progesterone is produced by the syncytiotrophoblast layer and regulates expression of SPP1 from the cytotrophoblast layer. SPP1 then binds to $\alpha_V\beta_3$ receptors in the syncytiotrophoblast layer and results in adhesion and communication between the two cell layers. It also helps to maintain the structural integrity of the chorionic villus by stimulating

changes in the morphology of the trophectoderm. This outer layer goes on to form the extraembryonic placental membranes during placentation (Johnson et al., 2000).

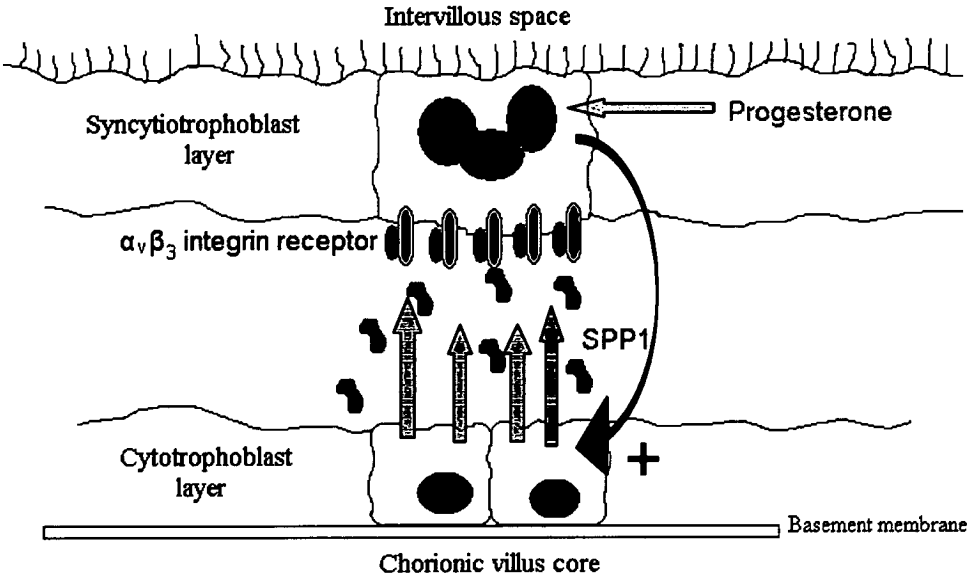


Figure 1-5 Trophectoderm (outer layer) of human blastocyst. Progesterone and/or interferon-tau stimulates release of SPP1 from cytotrophoblast layer which then binds to $\alpha_V\beta_3$ receptors in the syncytiotrophoblast layer and results in a paracrine feedback loop and adhesion and communication between the two cell layers Taken from (Omigbodun et al., 1997).

It is believed that the binding of SPP1 to its receptors also results in intracellular calcium oscillations, which allow the trophectoderm to signal to the uterine epithelium. This results in increased expression of SPP1 mRNA and subsequent secretion of the protein by the uterine epithelial cells (Johnson et al., 1999a; Omigbodun et al., 1997).

Other receptors apart from $\alpha_V\beta_3$ may be involved in the implantation process. As mentioned $\alpha_V\beta_1$ and $\alpha_V\beta_5$ heterodimers show affinity for the RGD binding motif of SPP1 (Hu et al., 1995) and the binding of SPP1 to $\alpha_4\beta_1$ receptors promotes leukocyte adhesion (Bayless et al., 1998). Bowen *et al.* (1997) investigated the porcine implantation sites of uterine epithelial and trophectoderm contact *in vivo* and

detected integrin subunits α_4, α_5, α_v, β_1 and β_3. They demonstrated very high levels of expression of these integrins during the porcine maternal recognition of pregnancy period (days 11-15). Porcine uterine luminal epithelia have been shown to increase their expression of α_4 and β_1 integrins in response to increased progesterone levels during the period of maternal recognition of pregnancy (Johnson et al., 1999a).

1.6. QTL for non-reproductive traits located on pig chromosome 8 (SSC8)

As discussed in section 1.4.1.4 several QTL controlling sow reproductive traits have been located on porcine chromosome 8. In order to refine these regions further and potentially use them within targeted marker assisted programs it is desirable to investigate whether any other QTL for other production traits also map to these regions, in order to avoid co-selecting undesirable alleles for additional traits of economic importance. Bidanel and Rothschild (2002) recently reviewed significant QTL regions identified in pigs for growth, fat deposition, meat quality and reproduction traits. Table 1-2 summarises the QTL for various traits identified on SSC8.

Table 1-2 Location of QTL for various production and behavioural traits on porcine chromosome 8

Trait	Location on SSC8	Genetic effect of QTL	Reference
Post weaning growth rate	P arm	Dominance effect of Large White alleles over Meishan alleles	De Koning *et al.*, 2001
On-test growth rate	P arm	Additive effect, Large White alleles display increased growth rate	Bidanel *et al.*, 2001
Back fat thickness	P arm and q arm	Favourable effect of Large White alleles	Rohrer, 2000 and Bidanel *et al.*, 2001
Carcass composition	Around centromere	Dominance effect of Meishan alleles over Berkshire alleles	Malek *et al.*, 2001
Carcass length	P arm	Additive effect of Large White alleles over Wild boar	Knott *et al.*, 1998
Bone/lean meat ratio	Q arm	Large White alleles superior over wild boar	Andersson-Eklund *et al.*, 1998
Meat quality – number of semi membranous muscles	P arm	Additive effect of Large alleles over Meishan alleles	Bidanel *et al.*, 2002 and Knott *et al.*, 1998
Immune capacity	Q arm	Additive effect of increased leukocyte counts associated with wild pig alleles versus Yorkshire alleles	Edfors-Lilja *et al.*, 1998
Exploratory activity towards a novel environmental stress	Q arm	Dominant effect Meishan alleles associated with higher activity than Large White alleles	Desautes *et al.*, 2002

1.7. Aims and objectives of current study

The aim of this study was to use a genomics and comparative mapping approach to identify chromosomal regions (QTL) and genes influencing female reproductive performance in pigs. The key biological resources used for this study were initially a series of three-generation Large White x Meishan pedigrees established for QTL-mapping and subsequently a collection of over 4000 Sygen commercial breeding sows with litter size records over several parities. As discussed earlier the Large White and Meishan sows deliver strikingly different reproductive performance. Therefore, the QTL mapping pedigrees provided an excellent resource with which to initiate the search for genes underlying this difference in performance. The data from Sygen provided a powerful tool for follow up association analyses (Figure 1-1). The scope of this study was limited to genes on porcine chromosome 8 (SSC8). The focus on SSC8 followed on from earlier studies in which QTL for reproductive traits had been mapped to SSC8 and by the mapping of a major gene (Booroola) to the homologous chromosome in sheep (OAR6) (see section 1.4.1.4).

The specific aims and objectives within the overall project are outlined below:

- To determine whether the Roslin Large White x Meishan pedigrees are segregating for QTL influencing female reproductive performance, and specifically for QTL on SSC8.
- To compare the map location and effects of such QTL with those previously mapped to SSC8 by other groups.
- To test the hypothesis that a QTL influencing ovulation rate and homologous to the Booroola gene is present in pigs. (The prior evidence of an ovulation rate QTL close to the neighbouring gene, *SPP1*, in pigs was inconclusive).
- To use a whole-genome radiation hybrid mapping panel to develop a gene map of SSC8.
- To align the gene map of SSC8 with homologous gene maps of other species.

- To identify candidate genes within any reproductive QTL identified, on the basis of known map locations in the pig and/or map locations predicted from comparative gene maps and/or physiological arguments.
- To scan such candidate genes for potential causal genetic variation
- To test candidate genetic variants by association analysis in the Sygen animals.

If any causal mutations or even QTL regions are identified to influence litter size, then it is ultimately hoped that targeted marker assisted selection programmes can be introduced. This will allow beneficial alleles from the more prolific Meishan breed to be selected for or introgressed into the European commercial breeds such as the Large White, without the introduction of alleles for the negative production characteristics of the Meishan. By increasing the size of a litter that an individual sow farrows, the size of the stock of females can be reduced and this will therefore produce a more efficient production system.

The understanding of the genetic control of ovulation rate or prenatal survival could also be linked to studies of reproductive physiology to help us to understand the underlying genetic problems associated with reduced fertility in humans. However, it must be noted that the reproductive mechanisms of pigs and humans are very different, in particular in the type of placenta formed and the number of offspring. With humans the blastocyst is invasive during implantation and the placenta forms a discoid morphology, whereas in pigs implantation is non-invasive and the placenta is diffuse (Johnson and Everitt, 1995). Another benefit of this project will be to the porcine genome mapping community, as it will result in an increase in the number of anonymous and gene-associated markers mapped to pig chromosome 8. Also by aligning the detailed gene map of SSC8 with maps of other species, it should help us to understand the level and order of conservation of genes across species.

Chapter Two

2. LOCATING QUANTITATIVE TRAIT LOCI (QTL) FOR FEMALE REPRODUCTIVE TRAITS ON PORCINE CHROMOSOME 8 (SSC8)

2.1. Introduction

The key porcine reproductive traits, including ovulation rate, prenatal survival and litter size, are expressed only in females and display low heritabilities (Bennett and Leymaster, 1989). Therefore, improvement of these traits in pigs by selective breeding has proved to be difficult. However, the substantial differences in reproductive performance between pig breeds indicate that there is useful genetic variation available for investigation. For example, the European Large White and the Chinese Meishan breeds show marked differences in fecundity. The Meishan is one of the most prolific pig breeds known (Haley et al., 1995), farrowing three to five more viable piglets per litter than European breeds. Thus, it is assumed that Meishan pigs are likely to carry alleles with the potential to enhance the reproductive performance of Western breeds (Haley and Lee, 1993).

As prolificacy is expressed in one sex, relatively late in life, it is a trait for which marker assisted selection (MAS) could be particularly beneficial (Avalos and Smith, 1987 and (Lande and Thompson, 1990). Information from genetic markers could be used to select males carrying desirable alleles for female reproductive performance and to select females without waiting for them to reach sexual maturity and have their first litter.

Two approaches have been pursued to identify genetic markers for reproduction traits. First, genome scans employing anonymous DNA markers have been used to identify quantitative trait loci (QTL) influencing these traits (e.g. Rathje *et al.*, 1997; Rohrer *et al.*, 1999 and (Wilkie et al., 1999). Secondly, the physiological candidate gene approach utilises polymorphisms within or close to genes known to have a role in reproduction in tests for associations with variation in fecundity (Rothschild *et al.*, 1996; Drogemuller *et al.*, 2001; Kmiec *et al.*, 2001; Jiang *et al.*, 2001 and van der Steen *et al.*, 1997).

Within this chapter the strategy has been to combine these two approaches, identifying QTL through a chromosome scan and testing genes identified as candidates on both positional and physiological grounds. I have focused on identifying QTL for reproductive performance present on pig chromosome eight (SSC8) only. Although earlier studies have provided good evidence for QTL on chromosome 8 influencing several reproductive traits, as yet none of these QTL have been confirmed across groups (Cassady *et al.*, 2001; Milan *et al.*, 1998; Rathje *et al.*, 1997; Rohrer *et al.*, 1999; Wilkie *et al.*, 1999 and Braunschweig *et al.*, 2001). Reported here is an independent confirmation of reproductive QTL on pig chromosome 8 and some candidate genes are eliminated on the basis of their map locations.

2.2. Materials and Methods

(Protocols for all solutions mentioned are detailed in appendix I)

2.2.1. Population structure

Three separate Meishan x Large White cross populations were developed at Roslin Institute over a period of eight years. These groups were defined as QTL 1, QTL 2.1 and QTL 2.2. The last two populations had a small number of grandparental individuals in common.

The purebred Meishan pigs were derived from an importation of 11 males and 21 females from the Jiadan county pedigree on the Lou Tang research farm in China in 1987 (Haley et al., 1992). The first animals used in this study were second-generation descendants of these imports. The purebred Large White pigs were from a control population derived from a broad sample of genotypes in 1982 (Cameron et al., 1988).

In all of the populations, F_1 reciprocal crosses were produced: Meishan (MS) male x Large White (LW) female and LW male x MS female and all of the F_0 animals were unrelated (Walling et al., 1998). From the subsequent F_1 generation, boars were mated to sows of a different grandparental pairing. The resulting F_2 female offspring were mated to one of a few selected purebred Large White boars and various reproductive traits recorded. In total there were 35 F_0 (13 males and 22 females) 94 F_1 (14 males and 80 females) and 220 F_2 females. Figure 2-1 shows the structure of the three-generation crosses. This diagram demonstrates clearly the segregation of alleles for the quantitative trait of coat colour in the F_2 generation. The genes controlling the quantitative reproductive traits such as litter size and ovulation rate will be inherited in a similar manner.

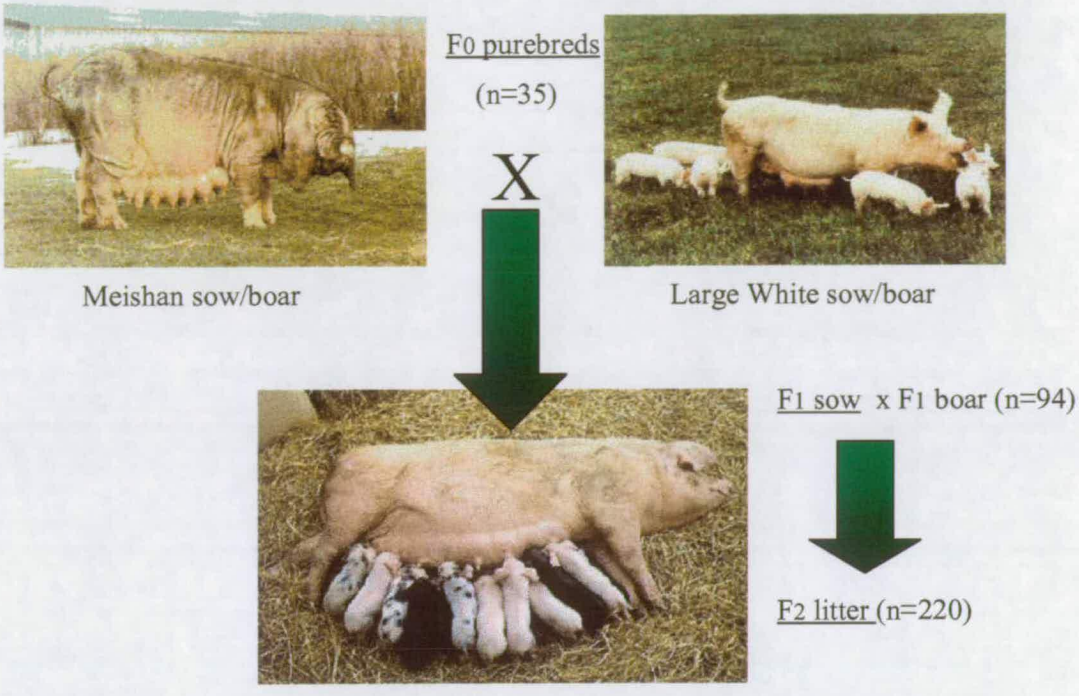


Figure 2-1 Diagrammatic representation of the three generation Meishan and Large White pedigree crosses (n=number of individuals in each generation).

The animals were a minimum 85 kg live weight at the start of each of the trials and they were reared indoors on *ad libitum* standard commercial growth rations until the time scheduled for first mating (for more detail see (Haley et al., 1992)).

Matings for each of the F_2 individuals, in the three different year groups, took place in two six-week periods. Gilts in the first age group were 8-11 months of age, corresponding to the animal's first parity. They were then re-mated at 13-17 months of age and individuals in this second age group mostly had their second parity; a few who had an unsuccessful first mating had their first litter at this later age. All sows were observed daily for signs of oestrus and were mated on the same day as detection to one of a few selected purebred Large White boars.

2.2.2. Phenotypic data recording

The trait data was recorded prior to the start of this Ph.D. and entered into the Roslin resSpecies database (http://www.resSpecies.org). At 5-20 days after mating the number of corpora lutea on the left and right ovaries was recorded by laparoscopy, this was used as an estimate of ovulation rate (Wildt et al., 1973). At the time of laparoscopy, the weight of the animal and the number of entire and partial teats on the right and left hand sides were also recorded. Some sows then returned to oestrus and if they were still within the first six-week mating period, were re-mated. For those animals, which were successfully re-mated in this manner, there is no record of the corresponding number of corpora lutea, because the mating occurred after the time of laparoscopy. These procedures were repeated for the same animals around five months later.

Prenatal survival was calculated as the number of piglets born divided by the total ovulation rate, for those individuals where their farrowing records corresponded to the ovulation rate recorded. Gestation length (days) was calculated as the difference between the age of the sow at mating and her age at farrowing. Analyses on the trait data were carried out within age groups and the variance within each of the traits calculated, to ensure there were no errors in the data recording and to confirm that the data displayed a distribution that was close to normal, an assumption of the QTL analysis. Individuals with incomplete records were removed from the analysis.

2.2.3. Genotyping DNA samples

At the end of each experimental trial all animals were culled and spleen samples were collected and stored at -70° C. DNA was prepared by standard procedures from these frozen spleen samples prior to the start of this Ph.D. project. The genotypes of the F_2 trait-recorded females, their F_1 parents and purebred grandparents were then determined for twenty polymorphic anonymous microsatellite and gene-associated markers, which were roughly evenly spaced across SSC8.

The size of the PCR product amplified by each of the primer pair for each microsatellite marker was determined and the markers labelled with one of three fluorescent dyes (hex, tet and fam) ensuring that the size range of markers labelled with the same dye did not overlap. This allows pooling of samples with different dyes and therefore optimises the efficiency of the method.

For each microsatellite marker the allelic DNA fragments were amplified from 75 ng aliquots of genomic DNA in 15 µl reaction volumes, containing 7.5 pmol of each primer (MWG Biotech AG, Ebersberg, Germany), 2.0 mM of each dATP, dTTP, dGTP and dCTP (Amersham Pharmacia Biotech inc, Little Chalfont, UK), 1.5 mM $MgCl_2$ in 1x PCR buffer (Roche Diagnostics, Mannheim, Germany) and 0.375 U *Taq* DNA polymerase (Roche Diagnostics, Mannheim, Germany). Where additional magnesium was necessary in the reaction mix, the required amount of double distilled water was substituted with $MgCl_2$ (25 mM) (Boehringer, Mannheim, Germany). The optimal polymerase chain reaction (PCR) conditions for each specific primer pair were determined and PCR amplifications performed using either a Hybaid™ Omnigene or Touchdown thermocycler. Appropriate dilutions of PCR products for microsatellite markers were pooled along with a 350 Tamra™ size standard (Applied Biosystems, Warrington, UK) and fractionated on 6 % polyacrylamide gels on an ABI 373 DNA sequencer (Applied Biosystems, Warrington, UK). The sizes of both allelic fragments for each individual animal and primer pair were estimated using the ABI GeneScan 2.1™ software (Applied Biosystems, Warrington, UK).

PCR-RFLP, Bi-PASA (Liu et al., 1997) and PCR-DSCP techniques were used to genotype the gene associated markers. Zhihua Jiang at the University of Guelph, Canada genotyped these ten gene associated markers. The PCR reactions were performed on ~50 ng of genomic DNA as templates in a final volume of 10 µl containing 3 pmol of each primer, 2.0 mM of each dATP, dTTP, dGTP and dCTP, 2.5mM $MgCl_2$, 50 mM KCI, 10 mM Tris HCI, 0.1 % Triton X-100 and 0.5 U of *Taq* polymerase. After denaturation at 94° C for 3 min, 30 amplification cycles were performed consisting of denaturation at 94° C for 30 sec, annealing at 61° C or 57° C

for 30 sec and extension at 72° C for 30 sec, followed by a further 5-min extension at 72° C.

For the PCR-RFLP assays, 5 µl PCR products were digested with 5 U of the desired restriction enzyme. Both PCR-RFLP and Bi-PASA products were then examined by electrophoresis on 1.5 % agarose gels with 1X TBE buffer and PCR products from PCR-DSCP assays were examined using 8 % acrylamide gels. The gels were stained with ethidium bromide and photographed.

All genotypes were entered into a resSpecies database (http://www.resSpecies.org) through a data submission tool that checks for inheritance errors in the data. Where errors were detected, the genotyping was repeated and the data point either corrected or removed if the problem could not be resolved.

2.2.4. Linkage map construction

MultiMap (Matise et al., 1994) is an expert system for linkage and radiation hybrid map construction and incorporates CRI-MAP version 2.4 (see http://compgen.rutgers.edu/multimap/crimap) as its core algorithm. These programs were used to determine the order of the markers in terms of the number of informative meioses and to build framework and comprehensive linkage maps.

CRI-MAP deduces any missing genotypes by calculating likelihood based on either known genotypes or from partial genotypes, where only one allele is known. The resulting linkage map was checked using the chrompic option in CRI-MAP to highlight potential genotyping errors involving double recombinants occurring within short map distances (i.e. less than 5 cM). Where these occurred, the genotyping was repeated and either corrected or removed and then the analysis repeated. The linkage map developed was then used for the QTL scan (see section 2.2.5).

2.2.5. QTL scan

The method used for QTL analysis of a three generation pedigree, derived from a cross between out bred lines involving the use of regression-based interval mapping (Haley et al., 1994), was effected using the QTL Express web interface (QTL Express: http://qtl.cap.ed.ac.uk (Seaton et al., 2002)). A "fixed QTL allele" model, in which genetically distinct founder lines, in this case Meishan and Large White pigs, were assumed to be fixed for alternative alleles at the QTL affecting the traits of interest, was used.

Each reproductive trait measured for animals in age group one and two was investigated individually for evidence of single and multiple QTL on chromosome 8. Where traits were measured individually on the left and right sides of the animal, only the sum of left and right values was investigated. In order to develop the model for each QTL analysis, the effect of the co-variates (age at mating, weight at laparoscopy, gestation length) on the individual reproductive traits measured, were investigated by the use of stepwise multiple regression analysis. For all QTL analyses experimental group (i.e. QTL 1, 2.1 and 2.2) was included as a fixed effect. Animals classed into age group two also had the parity of the sow (i.e. 1 or 2) included as a fixed effect.

Initially the additive (a) and dominant (d) coefficients were calculated for each marker. These values were then used to determine the information content (mean a+d) available to estimate the additive and dominance QTL effects for each of the individual markers. By using all the marker information simultaneously, the genetic information content (i.e. the amount of information available to determine the grandparental or breed origin of each allele) was estimated at 1 centiMorgan (cM) intervals along the whole chromosome (Seaton et al., 2002). The information content varies between 0 and 1, where a value of 1 means that the grandparental/breed origin is known with certainty.

Using ordinary least squares, the phenotypic values were regressed onto the additive and dominant coefficients to estimate the additive and dominance effects of putative

QTL at 1 cM intervals through the chromosome. The ratio of regression mean square to the residual mean square provided the variance (F) ratio test statistic, and the most likely QTL position was taken at the maximum value of F along the chromosome.

A single position permutation analysis was initially carried out, using 1000 permutations of the trait data, in order to determine the maximum nominal significance level of potential QTL for each trait. A chromosome-wide permutation analysis with 1000 permutations was then carried out to determine whether the QTL were significant (P <0.05) at the chromosome level, which is approximately equivalent to a suggestive QTL at the genome-wide level (Lander and Kruglyak, 1995).

The permutation test used for determining threshold values was first suggested by Fisher, (1935) and developed by Churchill and Doerge (1994). It involves the repeated shuffling of trait values, where each trait value is reassigned to a new individual whilst retaining the individual's genetic information and then a random sample of the test statistic (in this case the F value) is generated from an appropriate null distribution. Churchill and Doerge (1994) suggest that 1000 resamplings is sufficient to give a significance level of 5 % (P <0.05). Therefore where a significant association is observed between phenotypic data and genetic information at a particular location on the chromosome, then the significance threshold obtained can be used to ascertain whether the QTL is genuine i.e. the estimate of effect is greater than would be expected by chance.

For each trait investigated two genetic models were tested, a model where both the additive and dominance effects were fitted simultaneously and an additive, dominance and imprinting model. Where a significant F value for the QTL was detected, each of the genetic effects was tested for significance, to determine which type of effect or effects underpinned the QTL. (Obviously where no significant QTL were detected, none of the genetic effects would be significant). The estimates of the

genetics effects of each QTL were tested for statistical significance ($P < 0.05$) by the use of a two-tailed t-test.

2.2.6. Gene association analysis

Genetic markers within or close to several genes (*GNRHR, IBSP, STE, AREG, SPP1, SLIT2, QDPR, FGG* and *HD),* which includes some physiological candidate genes, were used to test for evidence of within breed marker associated variation for the reproductive traits of interest.

In the initial QTL analyses it was assumed that the two founder breeds were fixed for alternative alleles at the QTL. To the extent that the gene associated markers have allele frequencies that differed between the two founder breeds, an analysis looking for trait associations with a gene associated marker may be declared significant just because it explains some of the effect of a breed associated QTL. However, some of the gene-associated polymorphisms are not fixed for alternative alleles in the founder breeds and can therefore be used as an additional fixed effect in a QTL analysis to test for within breed variation at the QTL associated with the gene marker.

These tests for within breed variation were achieved by including the genotypes of each gene-associated marker as a fixed effect in the "fixed QTL allele" model as used previously. A resulting significant reduction in the residual mean square, i.e. a better fit of the data to the revised model, would indicate variation in the trait of interest associated with the gene marker over and above any due to the QTL. All of the gene-associated markers investigated were biallelic and individuals homozygous for one of the alleles were coded as 1, heterozygous individuals were coded as 2 and homozygous animals for the alternative allele were coded as 3. In theory similar analyses could also be implemented for each of the microsatellite markers. However, there are many more genotypic classes for multi-allelic markers such as microsatellites and the number of animals in each class would be too small.

An F test was used to determine whether the model with these gene-associated markers fitted as fixed effects gave a significantly better fit to the data than the initial

"fixed QTL allele" model. The value for the residual sum of squares calculated for the full model of interval mapping (the effect of the QTL, covariates and fixed effects) with the gene fitted as a fixed effect was compared to the equivalent value for the "fixed QTL allele" model. The following calculation was used:

$$F = \frac{(RSS_f - RSS_g)/(df_f - df_g)}{RSS_g / df_g}$$

Where RSS = residual sum of squares for the full model of interval mapping for the "fixed QTL allele" model ($_f$) and the gene fitted as fixed effect model ($_g$) and df = degrees of freedom (Numerator df = $df_f - df_g$ Denominator df = df_g).

For those models that revealed evidence of trait variation associated with the gene of interest, it was possible to investigate whether there was significant variation in the effect on the trait between any two genotypic classes for the gene of interest; for example whether homozygous animals differed from heterozygous animals at a particular locus. The significance of variation between genotype classes was tested using a two-tailed t-test.

2.3. Results

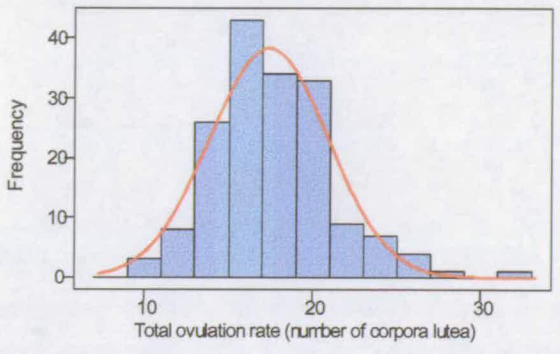
2.3.1. Analysis of phenotypic data

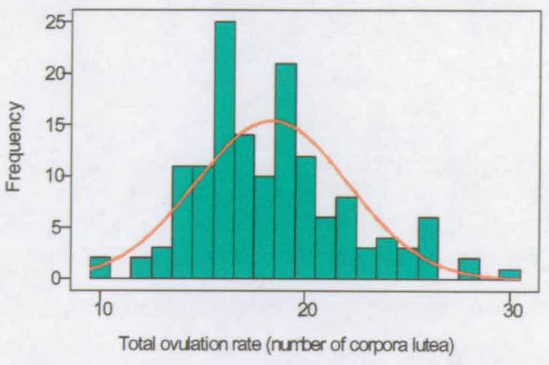
The variance within each of the traits recorded was initially analysed using simple descriptive statistics. This included the number of animals with data recorded (n), the minimum and maximum data values and the mean (± standard error of the mean) and standard deviation of the pooled data (Table 2-1). In order to investigate whether the distribution of the trait data was approximate to normal, a requirement of QTL analysis, histograms of the data were plotted against a normal curve using Minitab™ statistical software (release 13.32) (Figure 2-2). Each of the traits in both age groups, except for the trait of the number of partial teats, was shown to display a distribution close to normal. As the distribution of the number of partial teats was clearly skewed and as few non-functional teats were observed; it was decided that this trait would not be of interest for the QTL analysis and was not considered further.

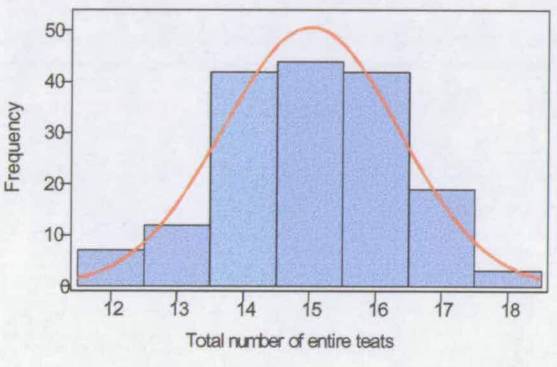
In addition, it was noted that the range of the total number of functional teats was different for the two age groups (Table 2-1). It would be assumed that the sows that had data in both groups would have the same number of teats. To verify this assumption, a Pearson correlation analysis was carried out for the 120 sows with teat number records in both age groups using Minitab™ statistical software (release 13.32) (Figure 2-3). For a few animals, the estimate across groups was different, but only by one or two teats. This is more than likely due to teats being classified as non-functional when they were functional and vice versa. Unfortunately it is impossible to be certain for these few animals what their true teat numbers were and it was decided that the data would still be analysed without any adjustments, to avoid introducing biases into the data set. Therefore teat number was only considered for the younger age group, as there were a larger number of animals with data in this group.

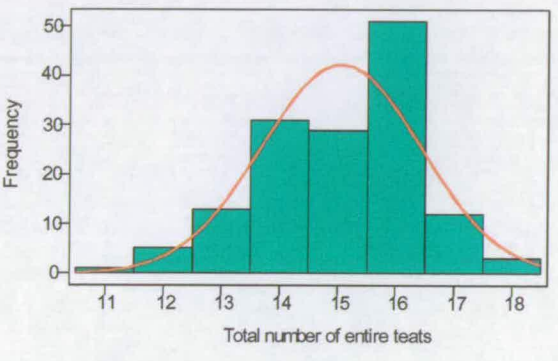
TABLE 2-1. Range, mean and standard deviation of values for each trait and covariate recorded. (n = numbers of animals). Note that calculations were only on those animals that had complete trait records. Within age group two animals, 26 sows had their first litter and 118 had their second litter.

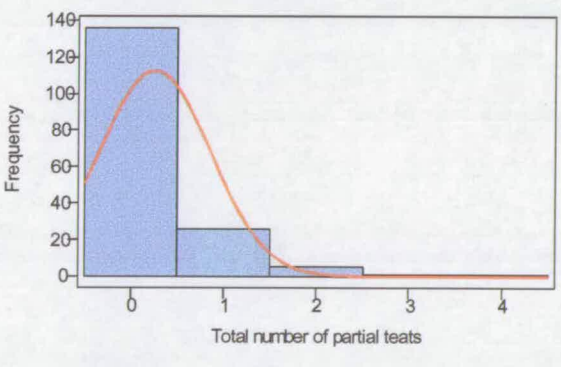
	Age group one (n=169)			Age group two (n=144)		
Traits Recorded	Range	Mean (±SEM)	SD	Range	Mean (±SEM)	SD
Ovulation rate on right ovary	0:15	7.74 (0.24)	3.06	3:18	8.97 (0.26)	3.14
Ovulation rate on left ovary	1:21	9.38 (0.26)	3.33	1:18	9.38 (0.25)	3.03
Total ovulation rate	9:31	17.12 (0.27)	3.52	10:30	18.35 (0.31)	3.73
Number of partial teats on right side	0:2	0.15 (0.03)	0.40	0:2	0.13 (0.04)	0.37
Number of partial teats on left side	0:2	0.12 (0.03)	0.35	0:2	0.10 (0.04)	0.33
Total number partial teats	0:4	0.25 (0.05)	0.60	0:4	0.23 (0.06)	0.60
Number functional teats on right side	6:9	7.52 (0.06)	0.76	6:9	7.53 (0.06)	0.72
Number functional teats on left side	6:9	7.49 (0.06)	0.80	5:9	7.51 (0.07)	0.80
Total number functional teats	12:18	15.01 (0.10)	1.33	11:18	15.04 (0.11)	1.36
Litter size	2:22	12.15 (0.28)	3.63	1:22	12.76 (0.28)	3.35
Prenatal survival	0.11:1.23	0.72 (0.01)	0.19	0.06:1.06	0.71 (0.01)	0.18
Covariates						
Age at mating (days)	248:357	300.74 (1.75)	22.77	402:559	491.62 (2.98)	35.75
Age at farrowing (days)	360:469	414.88 (1.75)	22.79	517:673	605.97 (2.99)	35.85
Gestation length (days)	108:119	114.14 (0.13)	1.64	111:118	114.35 (0.12)	1.45
Weight at laparoscopy (kg)	90:195	141.48 (1.58)	20.58	110:245	169.44 (2.30)	27.63

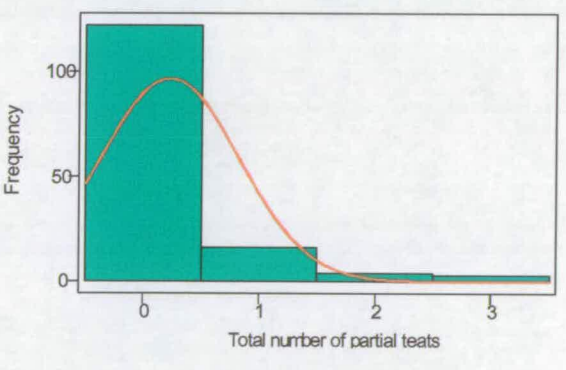

Frequency
0
10
20
30
40
10
20
30
Total ovulation rate (number of corpora lutea)

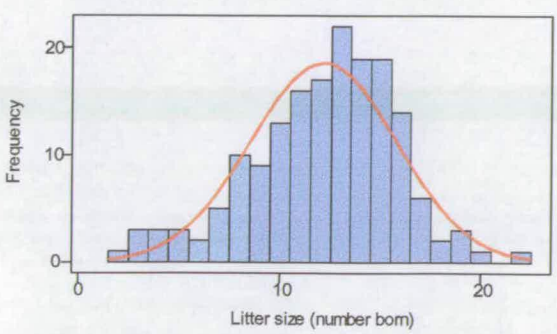

Frequency
0
5
10
15
20
25
10
20
30
Total ovulation rate (number of corpora lutea)

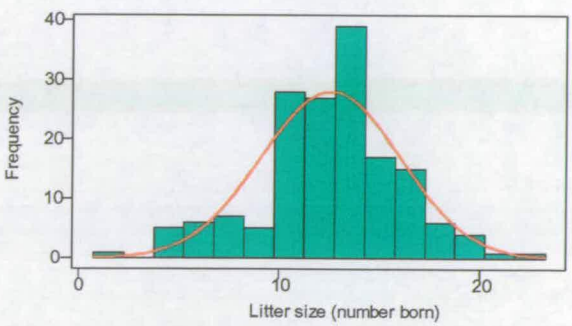

Frequency
0
10
20
30
40
50
12
13
14
15
16
17
18
Total number of entire teats

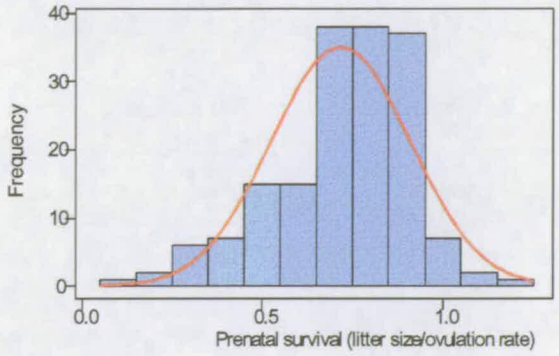

Frequency
0
10
20
30
40
50
11
12
13
14
15
16
17
18
Total number of entire teats

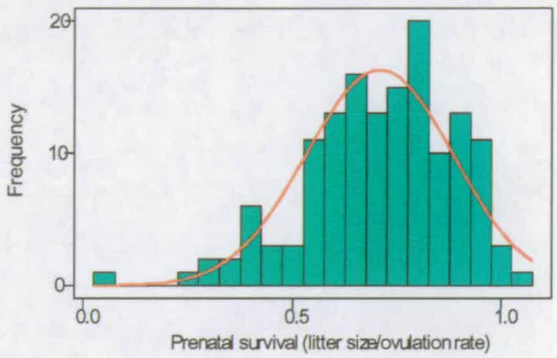

Frequency
0
20
40
60
80
100
120
140
0
1
2
3
4
Total number of partial teats

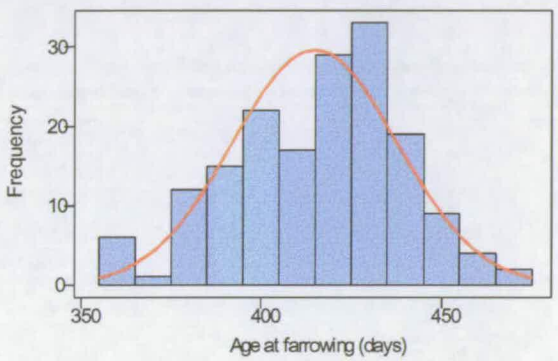

Frequency
0
50
100
0
1
2
3
Total number of partial teats

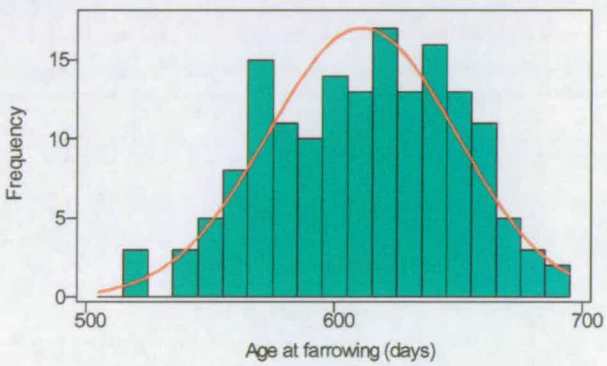

Frequency
0
10
20
0
10
20
Litter size (number born)

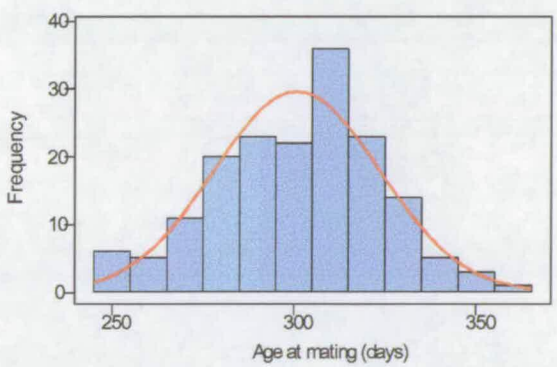

Frequency
0
10
20
30
40
0
10
20
Litter size (number born)

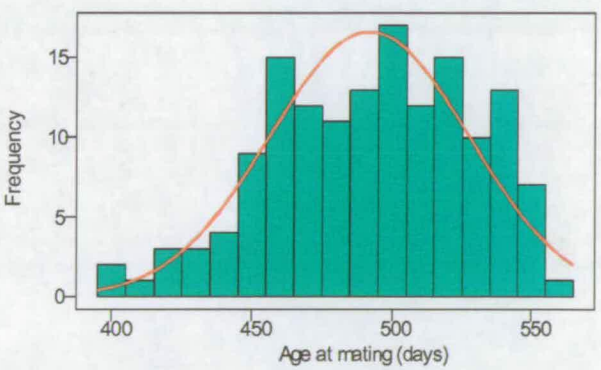

Frequency
0
10
20
30
40
0.0
0.5
1.0
Prenatal survival (litter size/ovulation rate)

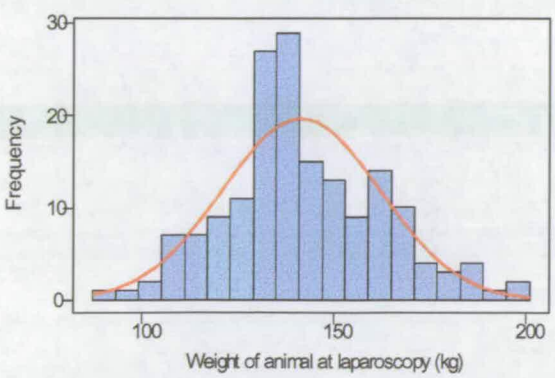

Frequency
0
10
20
0.0
0.5
1.0
Prenatal survival (litter size/ovulation rate)

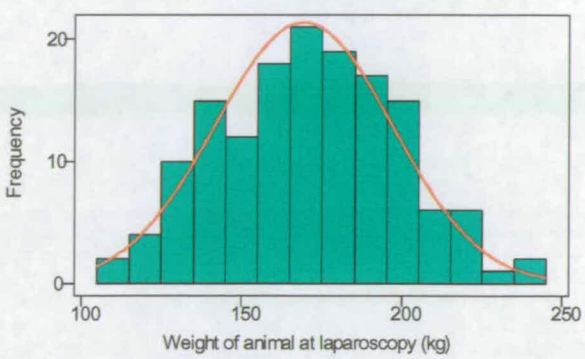

Frequency
0
10
20
30
350
400
450
Age at farrowing (days)


Frequency
0
5
10
15
500
600
700
Age at farrowing (days)


Frequency
0
10
20
30
40
250
300
350
Age at mating (days)


Frequency
0
5
10
15
400
450
500
550
Age at mating (days)


Frequency
0
10
20
30
100
150
200
Weight of animal at laparoscopy (kg)


Frequency
0
10
20
100
150
200
250
Weight of animal at laparoscopy (kg)

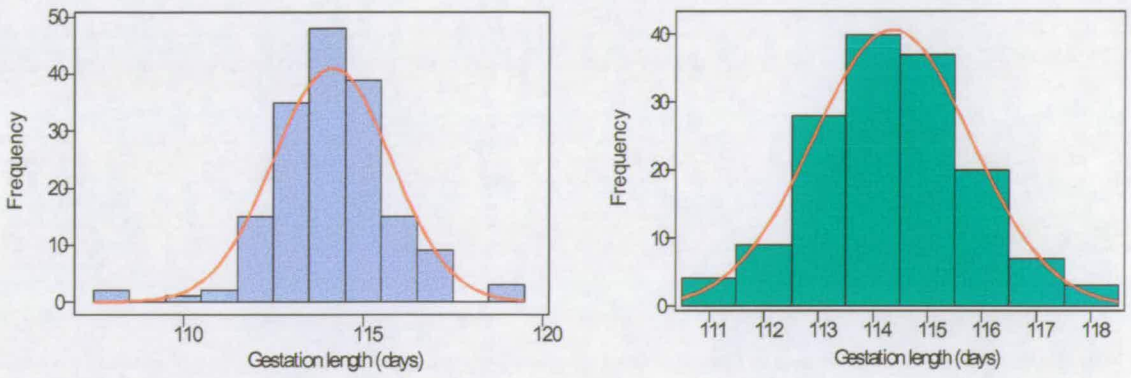
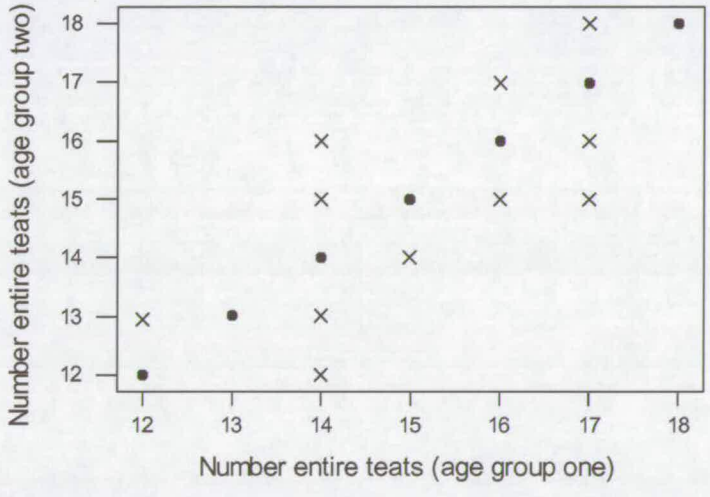


Figure 2-2 Histograms of trait data recorded shown along side a normal curve for animals in age group one (blue) and age group two (green).



Pearson correlation = 0.94 (P <0.001)

Figure 2-3 Correlation between the record of the number of teats for 120 sows with data in age group one and age group two. (x) single data point and (•) multiple data points.

The trait data were also checked for any extreme values. Outliers may have occurred as a result of incorrect data recording and these values would have to be removed. One individual in age group one was seen to have an ovulation rate of 31 ova, however this value is physiologically possible and therefore this data point was retained. In both age groups a total of three animals were recorded to have prenatal survival values greater than 1.0. Prenatal survival would not be expected to be greater than 1.00 (100 %). The most likely explanation is that the number of corpora lutea was underestimated at the time of laparoscopy. Another less likely explanation is that monozygotic twins were formed. This is where one zygote divides to form two genetically identical blastocysts (Johnson and Everitt, 1995). It was decided that these three values should be adjusted to 1.00, to allow for the error in the estimation of these data points.

Age at farrowing is strongly correlated with age at mating. The difference between the two variables is gestation length. Therefore in order to avoid unnecessary repetition, age at farrowing was not included as a variable in subsequent analyses.

2.3.1.1. Use of stepwise multiple regression to investigate the effect of the co-variates on the reproductive traits measured

In order to conduct the QTL analyses it was first necessary to determine which of the co-variates (weight at laparoscopy (kg), age at mating (days) and gestation length (days)) had a statistically significant effect (P <0.05) on each of the reproductive traits measured. The appropriate co-variate(s) could then be fitted into each QTL model. The performance of females in the three groups (section 2.2.1) was measured in different years and also at varying times of year. Therefore the animals will have been exposed to different external environmental factors; including day length, temperature and disease challenge. As a consequence, animals of identical genotype will have displayed different phenotypes due to the varying environmental inputs. It was also known that performance of the females differs between the first and subsequent parities. Therefore the group that each animal was in and the parity of the sow (i.e. 1 or 2) were included in all the regression analyses.

The stepwise regression analyses were carried out on Minitab™ statistical software (release 13.32). The analysis revealed that the weight of the animal was significantly correlated with the total ovulation rate for animals in age groups one (P <0.001) and two (P <0.01). Animals of a greater weight had an increased ovulation rate. For animals in age group one only, the gestation length was significantly associated with litter size and prenatal survival (P <0.05). Individuals with shorter gestation lengths had larger litters and higher levels of prenatal survival. Also increased weight of animals in age group one was shown to be associated with larger litter sizes (P <0.05). None of the other variables for animals in either age group had a significant effect on the other traits.

2.3.2. Genotyping of polymorphic markers on SSC8

Ten polymorphic microsatellite markers were typed as part of this project. The remaining ten gene-associated markers were typed by Zhihua Jiang (University of Guelph). The data for all twenty markers were combined for the QTL analysis. In order to visualise clearly the separate alleles present in each DNA sample for each of the ten microsatellite markers, the optimal PCR conditions for each primer pair was determined. This primer optimisation was carried out on the eight QTL1 F0 DNA samples. Table 2-2 shows the references to all 20 of the primer sequences and the PCR conditions used.

A total of 342 out of the 349 animals in all three populations had DNA available and were typed for each of the markers. Table 2-3 shows the allele frequencies and degree of heterozygosity for each marker. The percentage of undetermined genotypes for the seven microsatellite markers typed over all of the animals was low.

Table 2-2 Polymerase chain reaction (PCR) conditions used for each primer pair.

Marker		Reference
Anonymous DNA markers		
S0017	Microsatellite	(Coppieters et al., 1993)
S0178	Microsatellite	(Ellegren et al., 1994)[a]
S0225	Microsatellite	(Robic et al., 1994)[b]
SW7	Microsatellite	(Rohrer et al., 1994)
SW61	Microsatellite	(Rohrer et al., 1994)[c]
SW268	Microsatellite	(Rohrer et al., 1994)
SW905	Microsatellite	(Rohrer et al., 1994)
SW2410	Microsatellite	(Alexander et al., 1996)
SW2611	Microsatellite	(Alexander et al., 1996)
Gene-associated markers		
AREG-1	PCR-RFLP (*Sty*I)	(Jiang et al., 2002b)
FGG-2	Bi-PASA	(Jiang et al., 2002b)
IBSP-1	Bi-PASA	(Jiang et al., 2002a)
GNRHR-1	Bi-PASA	(Jiang et al., 2001)
GNRHR-2	Bi-PASA	(Jiang et al., 2001)
HD-1	PCR-RFLP (*Ava*II)	(Jiang et al., 2002a)
QDPR-1	PCR-RFLP (*Bfa*I)	(Jiang et al., 2002a)
SLIT2-1	PCR-RFLP (*Bst*NI)	(Jiang et al., 2002a)
SPP1-1	Microsatellite	(Moran, 1993)[d]
SPP1-5	Bi-PASA	(Jiang et al., 2002a)
STE-1	PCR-DSCP	(Jiang et al., 2002b)

The conditions for individual markers varied from the published data in the following way:

[a] 1.5 mM magnesium, denature cycle of 94° C 5 minutes, anneal phase of 30 cycles of 94° C, 58° C and 72° C each for 30 seconds and the extension cycle 72° C 5 minutes.

[b] 1.5 mM magnesium, denature cycle of 94° C 5 minutes, anneal phase of 30 cycles of 94° C, 55° C and 72° C each for 30 seconds and the extension cycle 72° C 5 minutes.

[c] 2.0 mM magnesium and 35 cycles of annealing at 60° C.

[d] 1.5mM magnesium, denature cycle of 95° C 5 minutes, 57° C 30 seconds and 72° C 1 minute. Anneal phase of 30 cycles of 94° C 45 seconds, 55° C 30 seconds and 72°C 45 seconds and extension cycle of 72° C 5 minutes.

Table 2-3 Allele frequencies, degree of heterozygosity and percentage of unknown genotypes for each of the ten microsatellite markers.

	S0178	*S0225*	*SPP1*	*SW61*	*SW2410*	*SW7*	*SW905*	*SW2611*	*SW268*	*SOO17*
Allele 1	0.50	0.55	0.44	0.54	0.40	0.60	0.29	0.18	0.07	0.52
Allele 2	0.37	0.13	0.23	0.12	0.30	0.31	0.26	0.07	0.07	0.28
Allele 3	0.09	0.12	0.20	0.08	0.19	0.09	0.18	0.28	0.21	0.10
Allele 4	0.04	0.10	0.06	0.07	0.11		0.18	0.43	0.15	0.10
Allele 5		0.07	0.06	0.06			0.09	0.04	0.17	
Allele 6		0.03	0.01	0.05					0.29	
Allele 7				0.04					0.04	
Allele 8				0.03						
Allele 9				0.01						
% heterozygosity	85.9 %	79.3 %	84.8 %	80.4 %	94.6 %	81.5 %	84.8 %	100%	100%	100 %
% unknown genotypes	1.2 %	1.2 %	4.4 %	2.9 %	0.9 %	0 %	5 %	76.3 % *	75.4 % *	75.4 % *

* The percentage of unknown genotypes is so high because these markers were only typed over the QTL1 population (n=91) and not all three populations (n=342)

2.3.3. Linkage map of pig chromosome 8

The complete sex averaged map produced of pig chromosome 8 was 139.3 cM in length. The order of markers and the length of the map were compared to the analogous linkage map of this chromosome produced at the USDA Meat Animal Research Center in Nebraska (Rohrer et al., 1996). Figure 2-4 shows that the two maps were of similar length, indicating a similar level of recombination between markers. Also the few markers, which were present on both maps, were in the same order.

2.3.4. Marker information content

The mean of the additive and dominant coefficients calculated for each marker, was used to determine the information content of the individual markers. The individual information content of each marker along the chromosome and the information content at 1 cM intervals using simultaneous marker analysis are shown in Figure 2-5. For the QTL analysis, simultaneous marker information was used and it can be seen that the information available to determine the QTL genotype at any one point along the chromosome was above 0.5, with the lowest area of information content on the p arm of the chromosome, around the marker *SW268*.

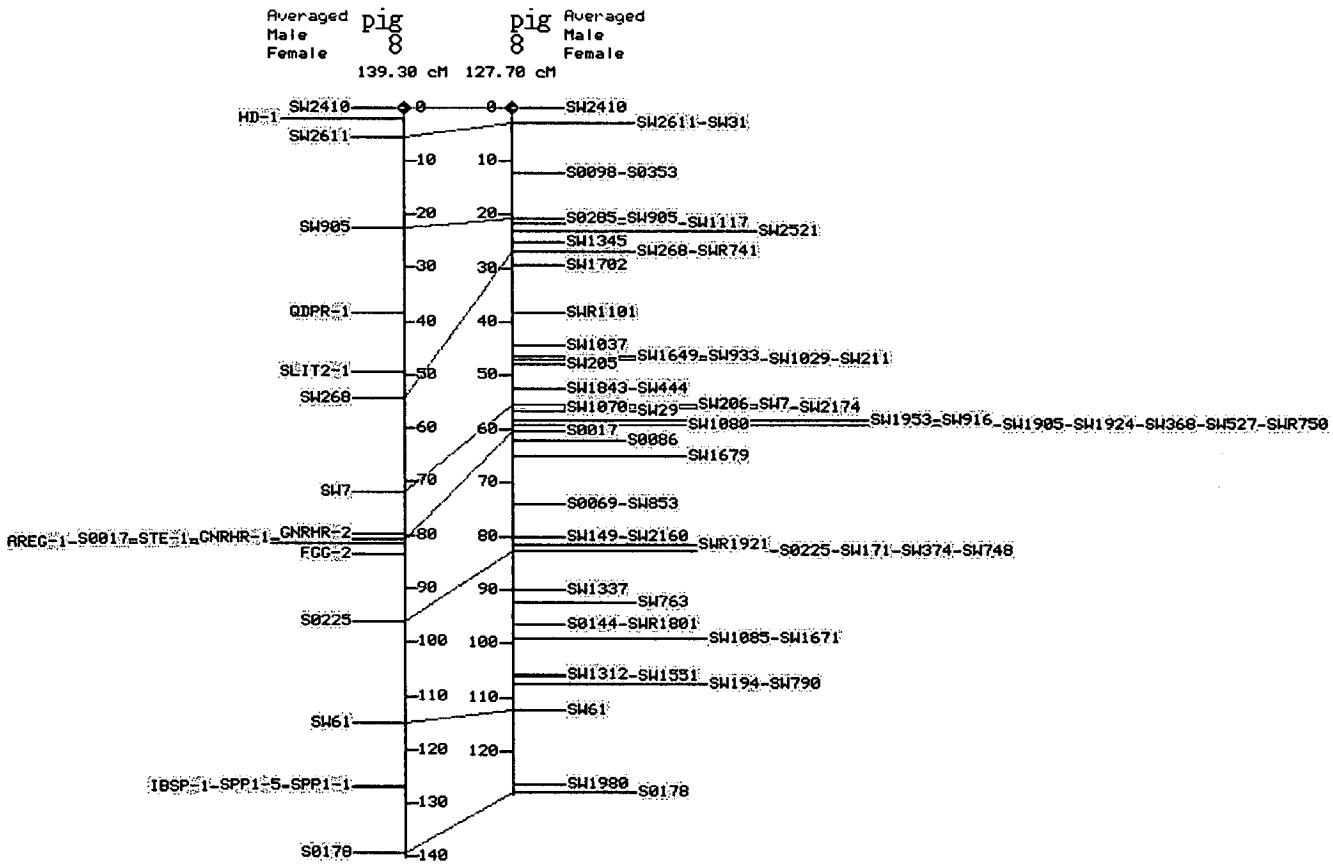


Figure 2-4 Pig chromosome 8 linkage maps (map produced by myself on left and at USDA-MARC on right). Microsatellite markers highlighted in blue, gene-associated markers in green. (SPP1-1 is a gene associated microsatellite).

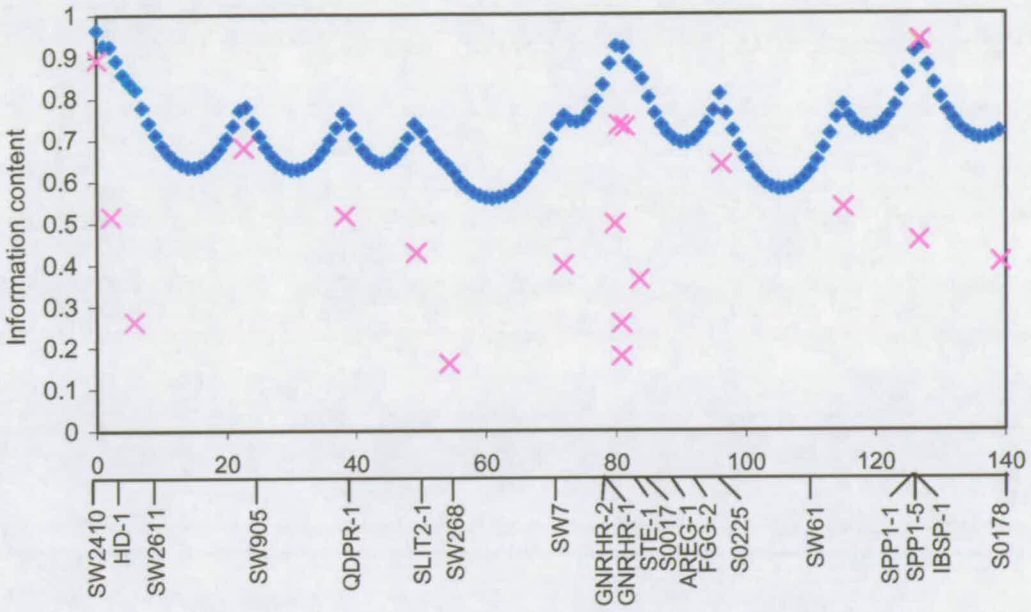


Figure 2-5 Individual information content of each marker along the chromosome (X) and also at 1 cM intervals using simultaneous marker information (◆). X-axis shows the relative position of the markers on the linkage map alongside the distances in cM on pig chromosome 8 (total length 139.3 cM).

2.3.5. QTL mapping

QTL analyses were carried out separately for each trait, for animals in both age groups, fitting the relevant covariates and fixed effects. The number of animals with genotype data and information available for each trait varied depending, for example, upon whether the ovulation rate recorded for an F2 individual corresponded to the subsequent litter that was born (see Table 2-4).

The estimated QTL locations and the corresponding significance levels are summarized in Table 2-4. The chromosome wide permutation analyses revealed two putative QTL, for prenatal survival (P <0.05) and number of teats (P <0.05) for animals in age group one only (i.e. first parity). A nominally significant QTL (P <0.01) for litter size was co-located with the prenatal survival QTL. Table 2-5 shows the estimates of the genetic effects for those QTL significant above the nominal level. The results are shown for the additive and dominance genetic model only, as the imprinting effect was shown not to be significant for all traits.

Table 2-4 The estimated QTL locations for all traits and the corresponding significance levels (N.S. = non significant (P >0.05 at nominal level)). Number of F2 animals = number of individuals with both genotype and phenotype records).

Trait	Number F2 Animals	Position on SSC8	F-ratio	Significance level (P value)
Age group one				
Total ovulation rate	179	2 cM	1.96	N.S.
Total number teats	193	49 cM	5.21	chromosome wide (P <0.05)
Litter size	152	127 cM	4.79	nominal (P <0.01)
Prenatal survival	152	125 cM	6.84	chromosome wide (P <0.05)
Age group two				
Total ovulation rate	153	139 cM	1.39	N.S.
Litter size	134	38 cM	2.09	N.S.
Prenatal survival	134	38 cM	2.70	N.S.

Table 2-5 Estimates of the genetic effects for those QTL significant above the nominal level. The additive effect was estimated as half the difference between the homozygotes for Large White versus Meishan alleles. Dominance effect was estimated as the deviation of the heterozygotes from the mean of the homozygotes. (Statistically significant genetic effects (P <0.05) are highlighted in bold).

Trait	Additive (±SE) Effect[a]	P value	Dominance (±SE) Effect	P value
Age group one				
Total number teats	-0.25 teats (0.14)	P >0.05	**+0.58 teats (0.20)**	***P* <0.01**
Litter size	-0.32 pigs (0.42)	P >0.1	**-1.80 pigs (0.60)**	***P* <0.01**
Prenatal survival	-0.02 (2%) (0.02)	P >0.1	**-0.11 (11%) (0.03)**	***P* <0.001**

[a]A negative additive effect indicates that the direction of effect is from the Meishan breed and a positive effect is from the Large White breed.

The interval mapping plots of each of the four traits for animals in both age groups are shown in Figure 2-6. The most significant QTL was seen for prenatal survival and the related trait of litter size for animals in age group one (Figure 2-6C). There was a clear peak in the F values calculated between markers *SW61* and *S0178*, close to the microsatellite marker 5' of the *SPP1* gene at the distal end of porcine chromosome 8q. It can be seen from the interval mapping plot that the QTL for teat number for animals in age group one, which is significant at the chromosome-wide level (P <0.05), is located near the *SLIT2* gene on porcine chromosome 8p (Figure 2-6B).

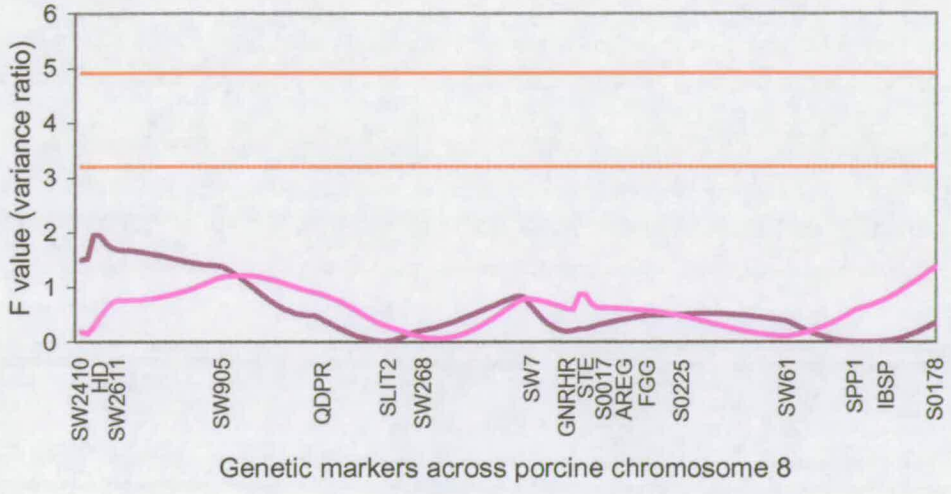
A two QTL model was tested for all of the traits, however teat number for animals in age group one was the only trait for which we found suggestive evidence for two linked QTL at 49 cM and 100 cM on SSC8. Interestingly the alleles at the two loci were seen to be acting in opposite directions; the genetic effects of these nominally significant QTL are summarized in Table 2-6.

Table 2-6 Estimates of genetic effects with two QTL model for teat number recorded for age group one animals

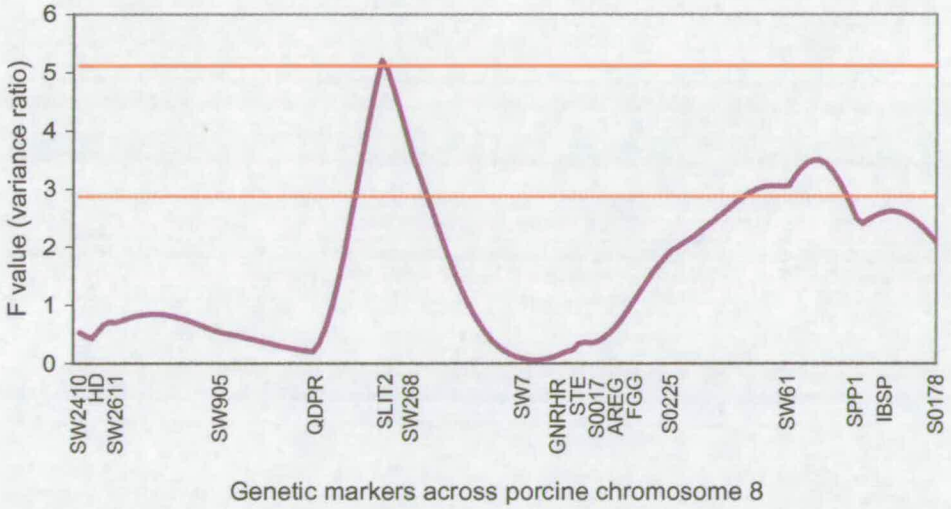
	Position on SSC8	Additive (±SE) Effect[a]	P value	Dominance (±SE) Effect	P value
QTL 1	49 cM	-0.34 teats (0.15)	P <0.01	+0.55 teats (0.20)	P <0.01
QTL 2	100 cM	+0.36 teats (0.15)	P <0.01	-0.16 teats (0.22)	N.S.

[a] A negative additive effect indicates that the direction of effect is from the Meishan breed and a positive effect is from the Large White.

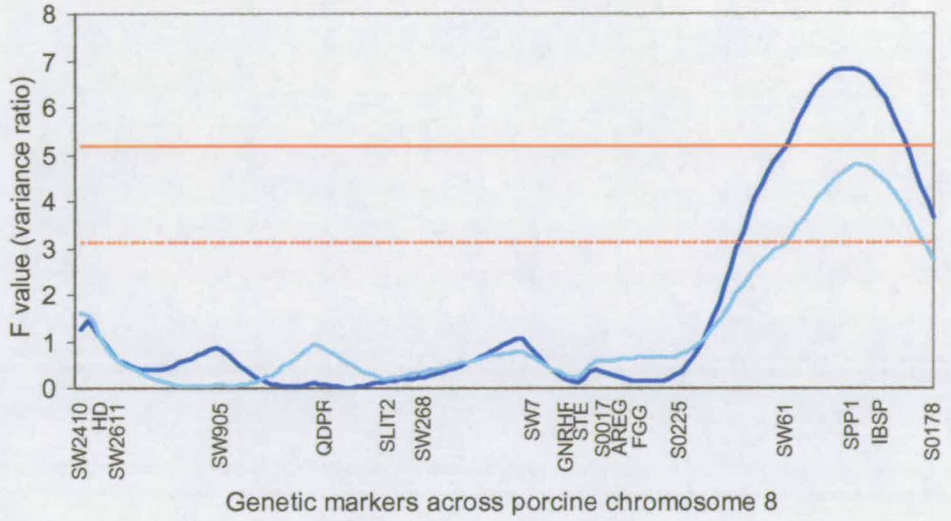
A. Ovulation rate for age group one (—, n=179) and age group two (—, n=153)



B. Teat number for age group one (—, n=193)



C. Prenatal survival (—, n=152) and litter size (—, n=152) for age group one.



D. Prenatal survival (—, n=152) and litter size (—, n=152) for age group two.

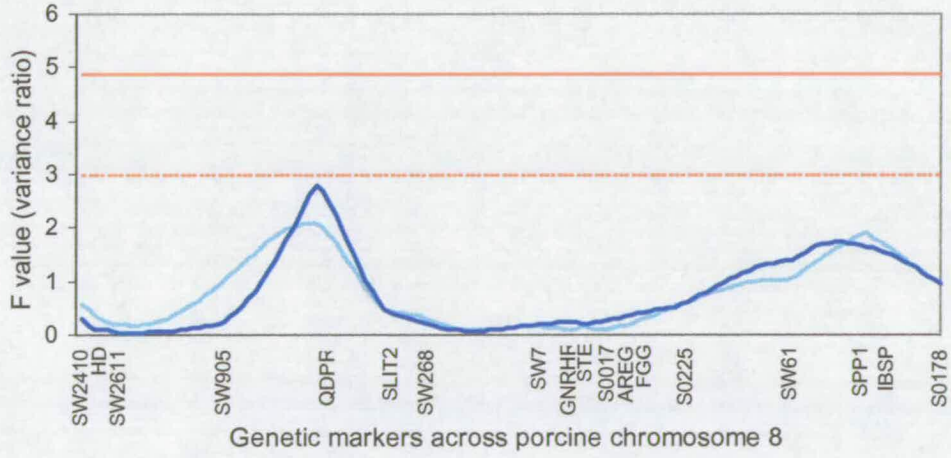
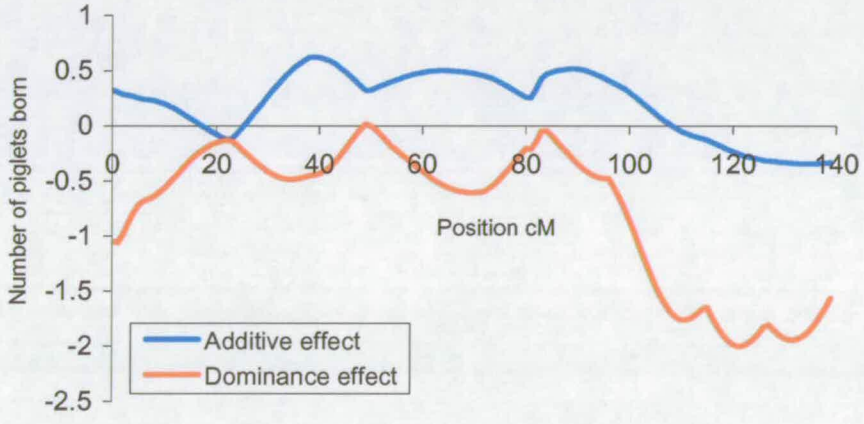


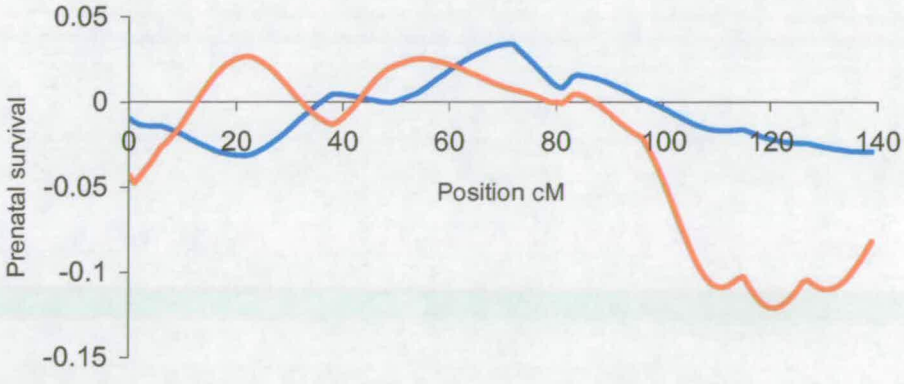
Figure 2-6 Interval mapping plots for ovulation rate (**A**), teat number (**B**), prenatal survival (**C/D**) and litter size (**C/D**) on chromosome 8 for animals in age groups one and two. Nominal significance level (- - - - -, P <0.05) and chromosome-wide significance level (——, P <0.05).

The additive and dominance effects across the chromosome for the significant QTL are shown in Figure 2-7. The plots clearly show that for prenatal survival and litter size (Figure 2-7A+B), the only area with a significant genetic effect on these traits is the dominance effect at around 120 cM. For the QTL for number of teats for age group one (Figure 2-7C), there is a clear additive effect from the Meishan breed and a negative dominance effect at around 50-60 cM. There also appears to be a second additive effect peak at around 100 cM, however the direction of effect in this region is from the Large White founders, as was observed with the two QTL model (Table 2-6).

A.



B.



C.

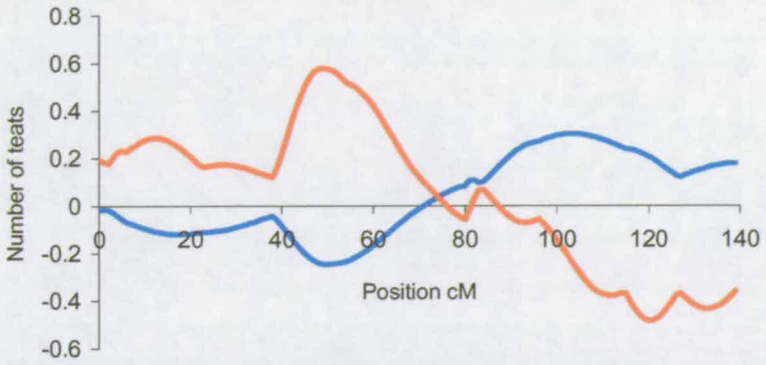


Figure 2-7 Additive and dominance effects across porcine chromosome 8 (x-axis) for those traits with significant QTL. A = Litter size B = Prenatal survival and C = Number of teats (all age group one animals)

2.3.6. Association analyses fitting candidate gene marker genotypes as fixed effects

As most of the biallelic gene loci were not fixed in the founder breeds, an assumption of the QTL analysis, they can be used to test for within breed variation. The allele frequencies in the founder breeds at each of these loci are shown in Table 2-7.

Table 2-7 The frequency of allele 1 in the founder breeds at each of the gene marker loci.

	Meishan purebred F0 animals	Large White purebred F0 animals
GNRHR-1	1.00	0.19
FGG-2	0.84	0.35
IBSP-1	1.00	0.00
GNRHR-2	0.97	0.39
STE-1	0.28	0.23
SPP1-5	1.00	0.00
AREG-1	1.00	0.04
HD-1	0.63	0.00
QDPR-1	0.94	0.31
SLIT2-1	0.69	0.12

For ovulation rate in age group one animals, *GNRHR-2* was the only genetic marker, which resulted in a significant improvement (P <0.01) in the fit of the model when added to the "fixed QTL allele" model. Under this extended model there is evidence for a QTL for ovulation rate (F-ratio 3.01) at around 3 cM. However this QTL is only significant at the nominal level (P <0.05). There was a second peak (F = 2.5) at around 80 cM, i.e. close to the *GNRHR* gene, however the F values calculated at this location were slightly lower than the nominal significance level (P >0.05) (Figure 2-8).

The effects of the ovulation rate QTL at position 3 cM was additive (P <0.05), with an estimate of +0.71 ova (± SE 0.33) per Large White allele. The effects of the putative QTL around 80 cM were also seen to be additive with the increasing allele inherited from the Large White founders. There is also a small region of additive effect from the Meishan founders at around 70 cM (Figure 2-9).

Indeed, a two QTL model revealed that there was evidence for two linked suggestive QTL for ovulation rate, one at 72 cM and one at 83 cM, acting in opposite directions. These two QTL locations are too close together to be resolved or indeed to be certain that they are genuine. The estimates of the genetic effects of these nominally significant QTL are summarised in Table 2-8. The t-test revealed that when the *GNRHR-2* genotypes were included as a component in a fixed QTL model, individuals homozygous for allele 1 had an increased estimate of 3.41 ± 1.47 ova (P <0.05) than individuals homozygous for allele 2.

Table 2-8 Estimates of genetic effects with two QTL model for ovulation rate with GNRHR-2 added to the "fixed QTL allele" model for age group one animals (a negative additive effect indicates that the effect is from the Meishan breed).

	Position on SSC8	Additive (±SE) Effect	P value	Dominance (±SE) Effect	P value
QTL 1	72 cM	-2.23 ova (0.71)	P <0.01	-1.55 ova (0.78)	N.S.
QTL 2	83 cM	+2.60 teats (0.73)	P <0.01	+0.29 teats (0.76)	N.S.

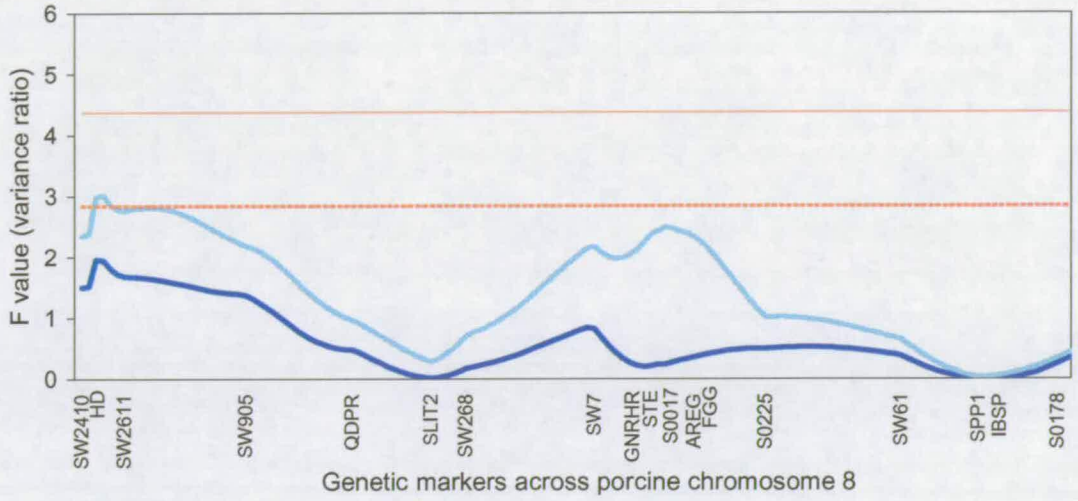


Figure 2-8 Interval mapping plots of ovulation rate for animals in age group one with "fixed QTL allele" model (——, n=179) and with *GNRHR-2* (——,n=179) genotypes fitted as fixed effects into the model. Nominal significance level (- - - - -, P <0.05) and chromosome-wide significance level (——, P <0.05).

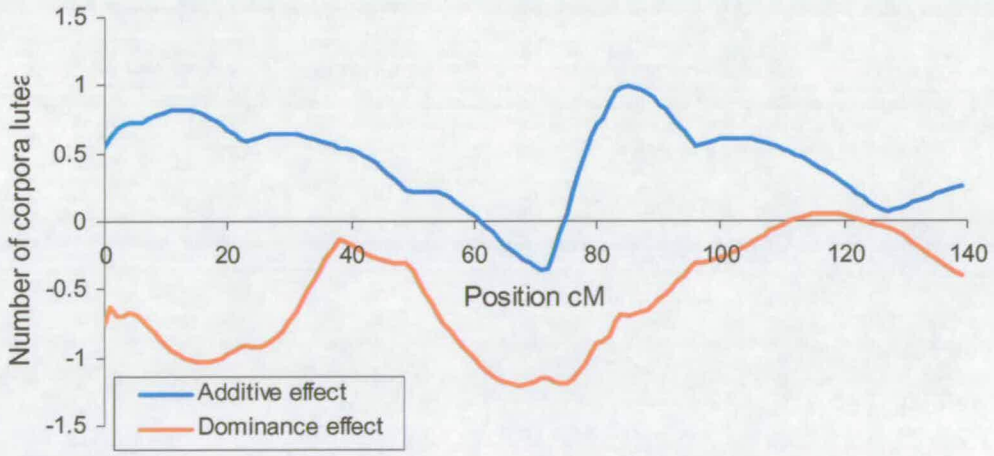


Figure 2-9 Ovulation rate genetic effects when *GNRHR-2* genotypes were fitted as fixed effects into the "fixed QTL allele" model (age group one sows).

Adding genetic markers associated with the genes *AREG* and *SLIT2* as fixed effects, resulted in significant improvements (P <0.005) in the model fit when compared to the "fixed QTL allele" model for teat number. Including *AREG* genotypes in the analysis model improves the evidence for the teat number QTL close to *SLIT2* on chromosome 8p, increasing the maximum F from 5.21 to 8.84 (Figure 2-10).

This estimate of the QTL is significant at the equivalent of a genome wide level (P <0.01 at chromosome-wide level). The additive effect of the QTL (P <0.01) was estimated as an increase in 0.49 teats (±SE 0.16) per copy of the Meishan allele. The dominance effect was also significant (P <0.01) with an estimate of +0.64 teats (±SE 0.20). Individuals who were homozygous for allele 2 at the *AREG* locus were significantly different (P <0.05) from individuals homozygous for allele 1, with an estimated increase of 0.68 teats (±SE 0.27).

The plot of the additive and dominance effects across the chromosome for the "fixed QTL allele" model and the model with *AREG* genotypes fitted is shown in Figure 2-11. This demonstrates that by including the genotypes at this locus as fixed effects, the effect at this particular gene locus on teat number from the Large White founders is removed. Therefore the estimate of the additive effect at the QTL increased, where the direction of the increasing alleles is from the Meishan founders. Therefore the evidence for the presence of the QTL at this location improved.

When the genotypes of the *SLIT2-1* locus were included as fixed effects, the QTL was no longer evident. It can be seen from Figure 2-10 that the peak F value is directly above the location of the *SLIT2* gene. It is highly likely that the QTL and the *SLIT2* genotypes are segregating together and therefore the genotypes are confounded with the breed origin.

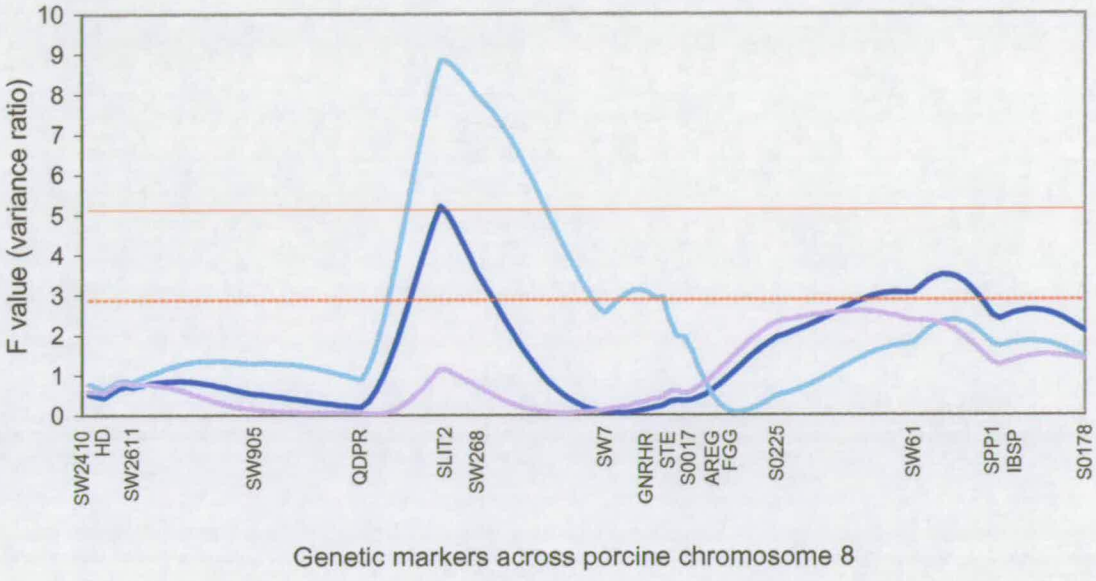


Figure 2-10 Interval mapping plots of teat number with "fixed QTL allele" model (, n=193) and with *AREG-1* (, n=193) and with *SLIT2-1* (, n=193) genotypes fitted as fixed effects into the model. Nominal significance level (, P <0.05) and chromosome-wide significance level (, P <0.05).

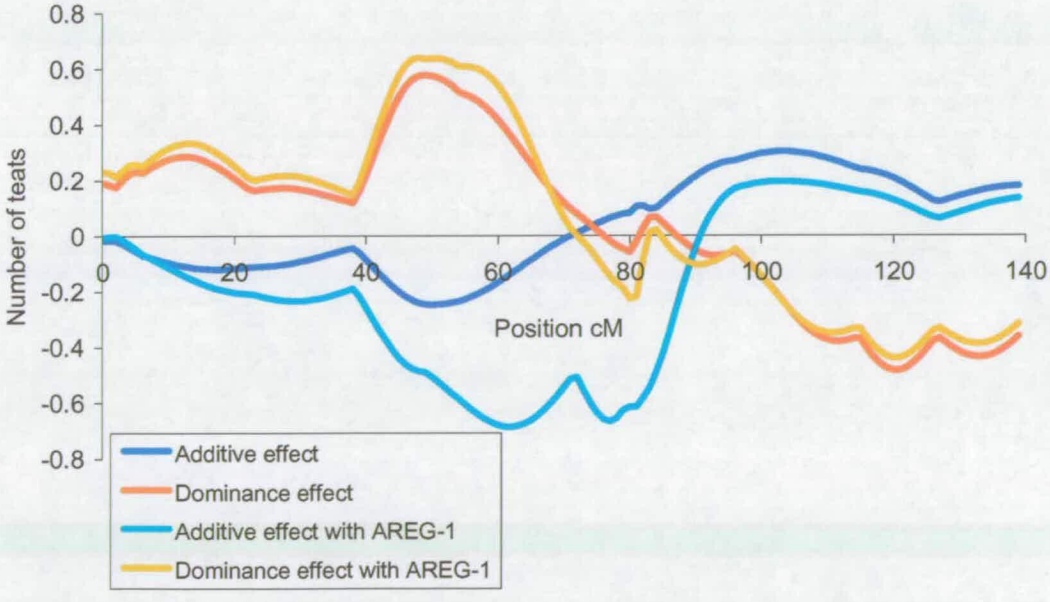


Figure 2-11 Teat number genetic effects when *AREG-1* genotypes were fitted as fixed effects into the "fixed QTL allele" model (age group one sows).

The genetic markers *SPP1-5* and *IBSP-1,* which are located close to the peaks for the litter size and prenatal survival QTL identified under the "fixed QTL allele" model for age group one sows (see Figure 2-6C), are fixed for alternative alleles in the founder breed (see Table 2-7). Thus, when these markers are added as fixed effects the fit of the model is not improved and due to the same confounding effects as with *SLIT2-1,* the evidence for the QTL is eliminated (No figure shown).

The only gene locus, which revealed a significant difference (P <0.05) between the two models tested for animals in age group two, was *QDPR-1* for prenatal survival (Figure 2-12). However no QTL were detected and consequently no significant genetic effects were estimated. The difference between the heterozygotes and 1,1 homozygotes at the *QDPR* locus was estimated to be + 0.1 (S.E. 0.03) or 10 %.

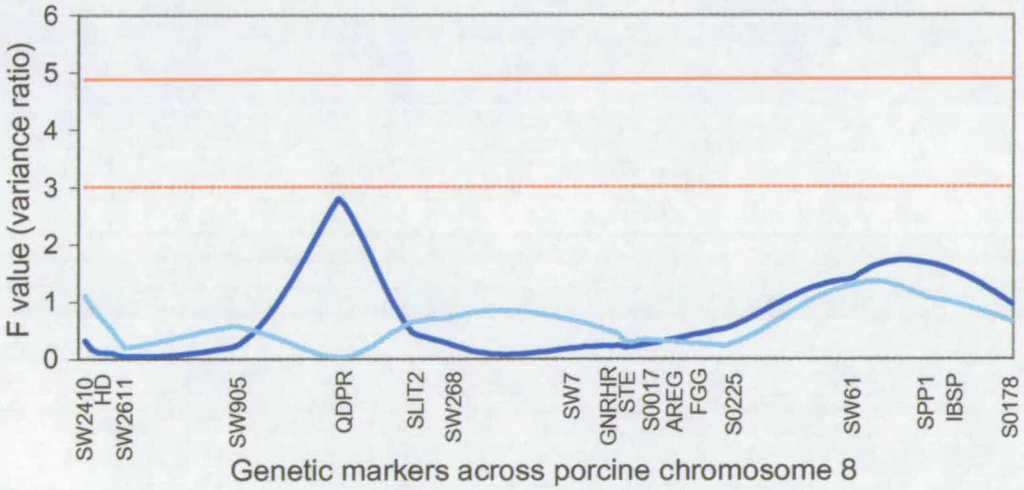


Figure 2-12 Interval mapping plots for prenatal survival for animals in age group two with "fixed QTL allele" model (, n=134) and with *QDPR-1* (, n=134) genotypes fitted as fixed effects into the model. Nominal significance level (, P <0.05) and chromosome-wide significance level (, P <0.05).

2.4. Discussion

From the analysis of the trait data it was concluded that there were no unusual outlier data points and that all of the values recorded for each of the traits were within normal physiological ranges. The animals in the first age group were between 8 and 12 months old at the time of laparoscopy and the range of weights was between 90 and 195 kg. The animals in the second age group were between 13 and 17 months old and the range of weights was between 110 and 245 kg, both these ranges of weights are normal for pigs of these ages. Interestingly, the mean number of functional teats was greater than the mean size of the litter. The range of gestation lengths in both age groups was 108-119 days, with a mean of 114 days. The gestation length of domestic pigs is typically 114 +/-6 days (Anderson, 1974).

The results of the QTL scan showed evidence for QTL towards the telomere of the q arm of porcine chromosome 8, controlling prenatal survival and litter size in young sows at first parity. The effects of these QTL suggest that they were both negative overdominant i.e. the heterozygotes show inferior performance to both classes of homozygotes. Although the additive effects were not significant, the beneficial alleles at this QTL appear to be from the Meishan breed. Such effects of the Meishan alleles at these QTL would be consistent with previous observations that the Meishan delivers its superior litter size through higher levels of prenatal survival for a given ovulation rate (Bidanel et al., 1989; Haley and Lee, 1993). However, evidence of positive overdominance had been reported in earlier studies (Haley and Lee, 1993).

There was also evidence for a QTL on SSC8 for teat number, the trait for which the greatest number of recorded animals was available. A QTL, significant at the chromosome-wide level, was detected on the p arm of chromosome 8 around the *SLIT2* locus with the increasing alleles coming from the Meishan founders. Interestingly when the genotypes at the *AREG* genetic marker were included as fixed effects in the model, the evidence for the QTL became stronger. The *AREG* polymorphism is close to fixation for alternative alleles in the Large White and Meishan founders (Table 2-7) with allele 2 only present in the Large White and

individuals homozygous for allele 2 at this locus were shown to have an increase estimate of teat number. The effects at the QTL (close to *SLIT2*) and around *AREG* are acting in opposite directions with the increasing allele associated with the Meishan and Large White respectively at these two locations. Thus, the improvement in the support for the QTL in the extended model is largely a result of fitting a two-locus model (i.e. a QTL plus *AREG* as a second locus) that reduces the interference from two QTL with opposing effects. By removing the "Large White effect" at the *AREG* locus the evidence for the QTL with an additive effect from the Meishan improves.

The *SLIT2* gene is of interest as it maps directly below the peak for the QTL with the positive alleles from the Meishan breed for age group one and two animals. *SLIT2* is a homologue of a Drosophila "*slit*" gene, which plays a critical role in central nervous system midline formation during embryogenesis (Rothberg et al., 1990). The human homologue of this gene is expressed in the spinal cord and it is believed that mammalian SLIT proteins may participate in the formation and maintenance of the nervous and endocrine systems by protein-protein interactions (Itoh et al., 1998). Therefore, there is no evidence to suggest that this gene plays a role in determining teat number.

Number of teats has not been investigated in most previous porcine reproductive QTL studies (Rathje *et al.*, 1997; Rohrer *et al.*, 1999; Wilkie *et al.*, 1999 and Braunschweig *et al.*, 2001). However evidence for a QTL affecting number of teats also on the p arm of SSC8 at the genome-wide significance level (P <0.05) has been reported (Cassady et al., 2001). Hirooka *et al.* (2001) found strong evidence for teat number QTL in a Meishan x Dutch cross on chromosomes 10 and 12, with the beneficial alleles from the Meishan breed and a QTL on chromosome 2 with a negative effect of the Meishan allele, but no evidence for QTL on SSC8. In addition Rohrer (2000) found significant evidence for a QTL for teat number in a Meishan x Large White cross also on porcine chromosome 10.

In agreement with the study by Rohrer *et al.* (1999), which also utilized a Meishan x Large White cross, I found some evidence for a QTL for ovulation rate at the p telomere of SSC8, when *GNRHR-2* genotypes were fitted as fixed effects in the QTL analysis. In both studies an additive increasing effect from the Large White breed was seen. These data provide insufficient evidence for this QTL to merit reporting the finding on its own. However, the estimated location of the QTL and the direction of the effect are consistent with the data of Rohrer *et al.* (1999). As the power to detect minor QTL is limited in this study and most previous studies of reproductive traits, it is important to report QTL for which the evidence is weak as confidence in such QTL can be strengthened by comparisons across studies. There would also be considerable benefit in analysing data pooled from several small to medium sized studies (cf. (Walling et al., 2000)).

I found no evidence for a QTL for ovulation rate at the telomere of the q arm, as reported by Rathje *et al.* (1997). However, when this research group included additional animals in a more comprehensive study, the previously reported QTL for ovulation rate at the telomeric end of SSC8q was not confirmed (Cassady et al., 2001). Cassady *et al.* (2001) described a QTL for age at puberty at SSC8q-ter, a trait that was not investigated in our study.

Wilkie *et al.* (1999) reported a putative QTL for ovulation rate around the centromere of SSC8, with a positive additive effect from the Yorkshire breed within the University of Illinois Meishan x Yorkshire Swine Resource Family. In a follow-up study, in which more markers were scored in these animals, the SSC8 centromeric QTL for number of corpora lutea was confirmed with increased confidence (Braunschweig et al., 2001). In my study, there was a region around the centromere of SSC8, which appeared to display an additive increasing effect from the Large White breed when *GNRHR-2* genotypes were added as fixed effects to the QTL model. Under this extended model there were two peaks at 3 cM ($F = 3.01$) and at 80 cM close to the *GNRHR* locus ($F = 2.5$), both displaying positive additive effects from the Large White breed.

It is recognized that the model in which it is assumed that the founder breeds are fixed for alternative QTL alleles is often an over simplification in outbred species such as pigs. Including within breed variation at the *GNRHR-2* marker represents a modest improvement in the sophistication of the model and provides a better fit for the ovulation rate data.

In an earlier study on the same animals used for this study, associations between the number of corpora lutea and *GNRHR* genotypes were tested (Jiang et al., 2001). The allele most prevalent in the Meishan breed for the polymorphism identified at position 1721 of the gene was shown to be associated with increased numbers of corpora lutea for animals at first parity. The genotypes for this polymorphism were included in this study as *GNRHR-2*, with the same allele coded as allele 1. Although allele 1 is present in both founder breeds it is much more abundant in the Meishan (Table 2-7) and as already mentioned 1,1 animals had an estimated 3.41 ±1.47 ova more than 2,2 homozygotes (P <0.05). Thus, the effects associated with *GNRHR* that maps close to the centromere appear to differ from the QTL reported for the Illinois population, in which the Meishan allele has a decreasing effect on the number of corpora lutea (Braunschweig et al., 2001; Wilkie et al., 1999).

As suggested by the two QTL model (Table 2-8) there does appear to be two QTL acting in opposite directions present in close proximity to one another around the centromere of SSC8. However the statistical evidence for two QTL so close together is too weak to be certain that they are genuine.

Milan *et al.* (1998) reported preliminary evidence for putative QTL for ovulation rate and litter size, with positive effects from the Meishan breed resulting in an increase of 1 or 2 ova or piglets, on chromosomes 7 and 8.

The statistical support for QTL, for all the traits examined across both age groups of sows, was more compelling for the younger age group. The main reason for this could be the loss of power resulting from the smaller number of F2 animals available in this second age group with trait data recorded.

One of the reasons for searching for reproductive QTL on pig chromosome 8 was that the pig homologue of the Booroola fecundity gene (*BMPR1B*) was predicted to, and is now known to map to the q arm of SSC8, relatively close to *SPP1,* but outside the confidence interval of the prenatal survival QTL (see Figure 3-8). Alleles at the *BMPR1B* locus are known to improve litter size in sheep through increases in ovulation rate (Souza et al., 2001; Wilson et al., 2001). However, as I found no evidence for an ovulation rate QTL on SSC8q where pig *BMPR1B* maps and as the litter size QTL effects appear to be attributable to improvements in embryo survival, it seems unlikely that *BMPR1B* is the gene responsible for the pig litter size QTL reported here.

Interestingly though, variation in a microsatellite repeat 5' of the *SPP1* gene was previously shown to be associated with an increase in litter size in a Meishan x Large White cross (van der Steen et al., 1997). This marker lies within the 95 % confidence intervals for the litter size and embryo survival QTL. More recently, Korin-Kossakowska *et al* reported associations between the presence of a SINE in the *SPP1* gene (Knoll et al., 1999) and litter size of the second and subsequent parities for 519 sows from a commercial Polish line (Korwin-Kossakowska et al., 2002).

SPP1, also known as osteopontin, is a physiological as well as a positional candidate gene. Studies of this gene have revealed that it is expressed in a variety of tissues, including the epithelial cells of the endometrium and the metrial gland cells of the decidua within the uterus, the placenta and the invading trophoblast, during the defined window of receptivity of early pregnancy of several mammalian species (Johnson et al., 1999a; Nomura et al., 1988). These studies indicate that this gene has an important role to play in embryo implantation and placentation.

The SPP1 protein is secreted into the uterine lumen during early pregnancy in humans and ewes and binds to integrin heterodimer receptors expressed on luminal epithelial cells, in response to increased levels of progesterone (Omigbodun et al., 1997). This induces adhesion between the luminal epithelium of the endometrium

and the trophectoderm of the blastocyst and triggers the cascade of molecular events leading to successful implantation and placental function (Fazleabas et al., 1997; Johnson et al., 2001). The same process is believed to occur in pigs (Garlow et al., 2002).

Though the *SPP1* gene is a candidate for the litter size and prenatal survival traits based on positional and physiological arguments, the confidence intervals for the QTL identified in this study are large, harbouring hundreds of genes. Further investigation of the QTL regions is therefore required in order to identify additional physiological candidate genes.

Chapter Three

3. RADIATION HYBRID MAPPING OF MULTIPLE MARKERS ON PORCINE CHROMOSOME 8

3.1. Introduction

When this project was initiated the number of genes mapped to pig chromosome 8 was limited and fewer still had been ordered along the chromosome. It was therefore necessary to pursue a comparative positional candidate gene strategy to identify trait genes within the QTL regions identified (the details of this strategy are described in chapter 4). In order to construct a detailed gene map of chromosome 8 the technique of radiation hybrid mapping was used.

Radiation hybrid (RH) mapping is a somatic cell technique used for ordering markers along chromosomes and estimating distances between them (Cox et al., 1990). It is a flexible and efficient method, which can be used to develop detailed physical maps. The main advantage of RH mapping is that the markers used are all unique Sequence Tagged Sites (STS) (Green et al., 1991), which are not required to be polymorphic (Hawken et al., 1999). This therefore allows the use of a wider spectrum of DNA markers than is available for meiotic mapping (Cox et al., 1990). The STS markers are easy to develop and are often linked to known genes within the genome.

Radiation hybrid panels are created by the use of irradiation and fusion gene transfer (IFGT), a technique originally developed by Goss and Harris (1975) who demonstrated that chromosomal fragments generated by the lethal irradiation of human cells could be rescued by fusion to rodent cells. Due to the lack of genetic markers available at the time this technique was not widely used. Cox *et al.* (1990) revived the technique, but this time panels were developed for single chromosomes only. The main disadvantage of this is that for example twenty-four separate panels would be required to map markers to the whole of the human genome. Therefore the technique was adapted for use with whole genome mapping by Walter *et al.* (1994), where a single panel represents the genome of a species.

Whole genome radiation hybrid panels are now available for several mammalian species including human (Gyapay et al., 1996), cattle (Womack et al., 1997), pig (Yerle *et al.*, 1998 and (Hawken et al., 1999)), mouse (Avner *et al.*, 2001 and (Hudson et al., 2001) and rat (Bihoreau et al., 2001).

The technique of IFGT involves irradiation with a predetermined dose of X-rays of a diploid fibroblast cell line of the species of interest. This cell line contains a specific selectable marker and once irradiated, the resulting chromosome fragments are fused with a recipient somatic cell line deficient in the selectable marker. The resulting hybrid clones containing one or many donor fragments are then selected. Intact donor cells line are lethally irradiated and whole recipient cell line are selected against (McCarthy, 1996).

Figure 3-1 shows an example of IFGT for the creation of a porcine–Chinese hamster radiation hybrid panel. Walter *et al.* (1994) showed that the retention of loci far away from the selected donor thymidine kinase gene were also retained at a relatively high frequency at around 20.9 % of all the hybrid cell lines tested. They concluded that where marker retention frequencies were between 20 and 50 %, 100 individual hybrids would provide sufficient coverage to produce a high resolution map of any chromosome in the whole genome (Walter et al., 1994).

In the same way as linkage map construction and in contrast to many physical mapping methods, radiation hybrid mapping determines statistically derived estimates for likelihoods of several different marker orders and the distances between them (Boehnke et al., 1991). The "best" map does not necessarily represent the actual order of the markers on the chromosome (Cox et al., 1990). The distance between two markers is measured as the frequency of irradiation-induced breakage between them and the likelihood of the estimate of that distance is measured by a LOD score (Gyapay et al., 1996).

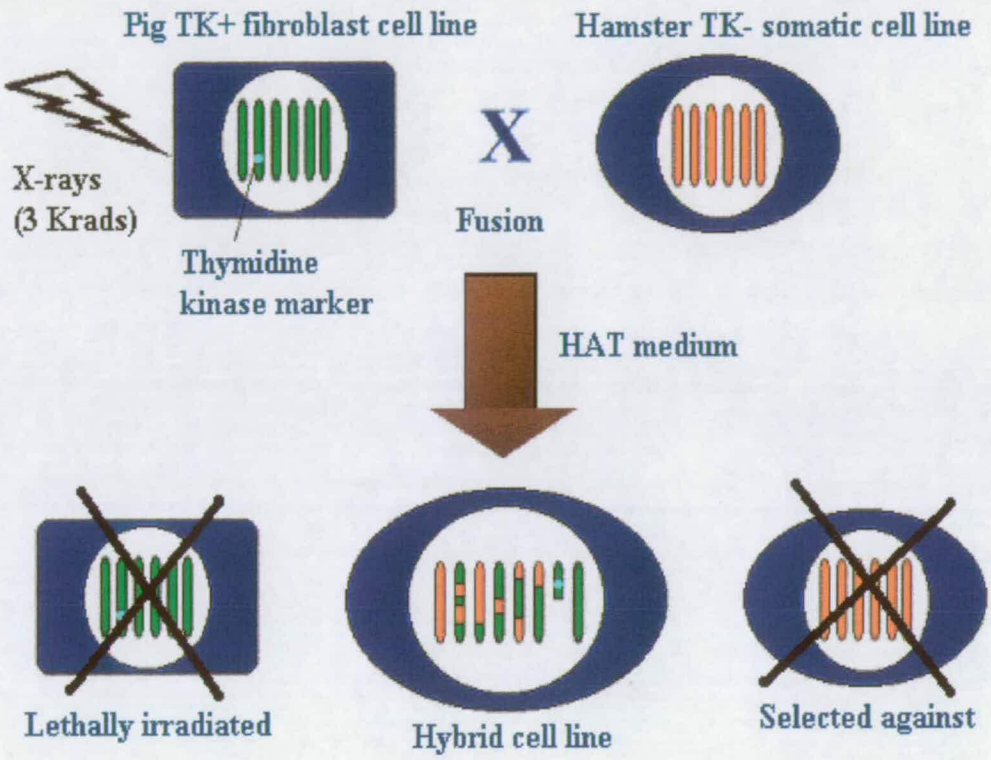


Figure 3-1 Schematic of irradiation and fusion gene transfer for the creation of a porcine-Chinese hamster radiation hybrid panel. The selectable marker within the porcine genome is thymidine kinase. Hypoxanthine Amino Thymidine (HAT) medium is used to select against whole TK- hamster cells and whole porcine cells are lethally irradiated by the X-rays. The resulting radiation hybrid cell line contains porcine chromosomal fragments on a hamster genomic background. A sufficient number of hybrid cell lines are chosen in order to ensure the whole of the porcine genome is represented.

The closer together two markers are, the less likely that there will be an irradiation induced break between them (Lunetta et al., 1995) and they will be co-retained in several hybrid clones more often than more distant markers (Ben-Dor et al., 2000) (see Figure 3-2). A number of algorithms have been developed to construct maps based on the pattern of marker retention. These maps can then be used to estimate the pair wise physical distance between individual markers (Schiex et al., 2001).

Linkage maps rely on recombination events between markers and around 1 % meiotic recombination or 1 cM corresponds to about 1 million base pairs (Mb) of DNA. In contrast, radiation hybrid mapping has been used to order markers within the human genome at a resolution of 50 kb (Lunetta et al., 1996). Low-resolution genetic maps can be used to identify QTL regions, however high resolution radiation hybrid or physical maps are needed to identify putative candidate genes within these regions.

Furthermore radiation induced breakage is believed to be random, i.e. independent of the structure of the DNA sequence (Slonim et al., 1997), whereas recombination occurs non-randomly throughout the genome and this would therefore bias the estimated distances between markers. For example a large genetic distance can be resolved as a small physical distance (Hawken et al., 1999).

A distance of 1 centiRay is equivalent to a 1 % frequency of breakage between two markers (Cox et al., 1990). An advantage of radiation hybrid (RH) mapping is that maps can be constructed at different levels of resolution by varying the radiation dose with which the donor cells are treated. Although increasing this dose will increase the resolution of the panel, the absolute relationship between dose, break frequency and resolution has not been defined in mathematical terms. The effort involved in determining such relationships given the number of potential variables and sources of experimental noise would not be justified. Rather groups developing RH panels have adopted a pragmatic approach.

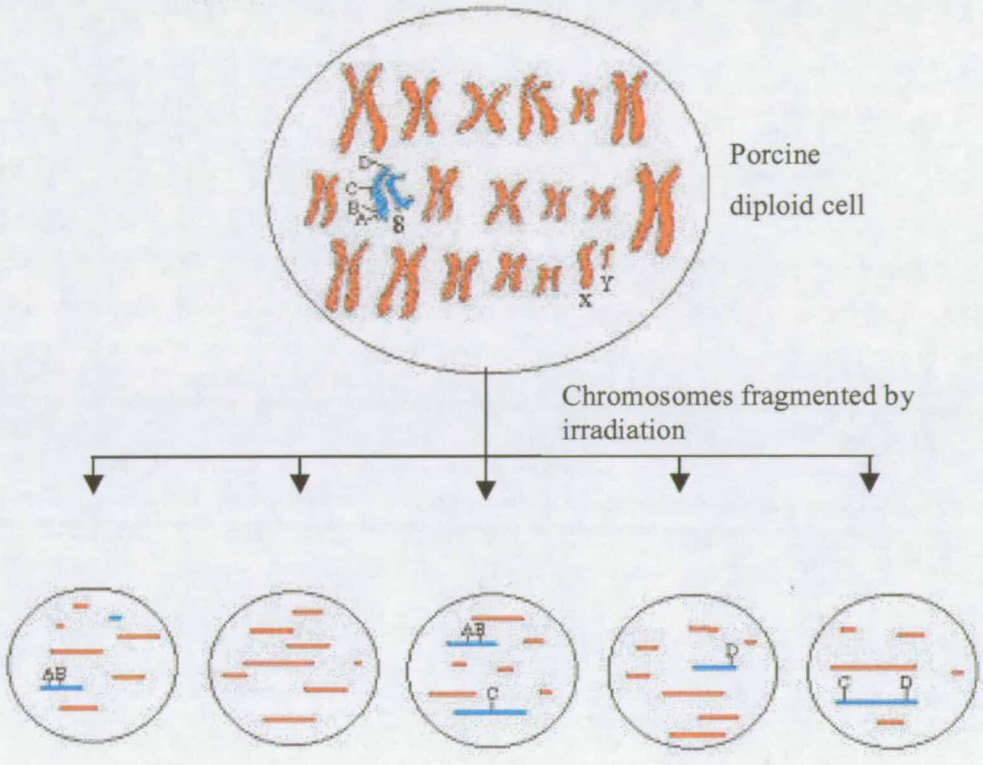


Figure 3-2 Schematic demonstrating how markers closer together on for example porcine chromosome 8 (e.g. markers A and B) will be co-retained more often than markers further apart (e.g. markers C and D and B and C). Note that for clarity only porcine fragments within the hybrid clones are shown. Generally hybrids contain porcine fragments integrated into hamster chromosomes. It is possible to find chromosomes of entirely porcine origin however these will be composed of pig fragments from different chromosomes.

Examples of the effect of increasing the radiation dose include the construction of a human RH map using 3000 Rads of irradiation, where 1 cR_{3000} was seen to be roughly equivalent to 300 kb (Gyapay et al., 1996) and in contrast a human map built with 10000 Rads, where 1 cR_{10000} was seen to be equivalent to a physical distance of around 29 kb (Stewart et al., 1997).

The main aim of this chapter has been to map additional genes within the teat number and the prenatal survival/litter size QTL regions identified in chapter 2 for age group one sows (Figure 2-6B and C), than had already been mapped to the meiotic linkage map (Figure 2-4). The strategy was to build a high-resolution framework map of several microsatellite markers on SSC8 and add onto it markers for porcine genes already known to map to chromosome 8. The plan was to then identify and map comparative positional candidate genes derived from the search of human chromosome 4 genes (Tables 4-1 and 4-2), to the porcine chromosome 8 QTL regions. This would give an improved alignment of the maps from the two species and improve the search for physiological candidate genes in the regions of interest.

3.2. Materials and methods

(Protocols for all solutions mentioned are detailed in appendix I)

3.2.1. Background of the radiation hybrid panel used

The pig-hamster whole genome radiation hybrid panel was established in a collaboration between Roslin Institute and the Department of Genetics, University of Cambridge. The donor cell line, which was a primary fibroblast line established from a male F1 Large White x Meishan pig, was exposed to 3,000 rads of X-rays and the fragments of porcine chromosomes fused to a thymidine kinase deficient (tk-) hamster cell line (A23).

Although DNA was prepared from 101 hybrids, a subset of 94 hybrids was assembled with pig and hamster control DNA, in a 96-well format. The technique used to produce the hybrid clones of porcine DNA on a hamster background is known as irradiation and fusion gene transfer (see Figure 3-1).

3.2.2. Primer design

All primer sequences for the DNA markers to be mapped over the radiation hybrid panel were either obtained from the literature or designed using the web based program Primer3 (Rozen and Skaletsky, 2000) (http://www-genome.wi.mit.edu/genome_software/other/primer3.html). The region of DNA for amplification was imported into the program and left (forward) and right (reverse) primers were designed accordingly. The primers were then synthesised by MWG Biotech AG, Ebersberg, Germany and stored at -20° C at a concentration of 100 pmol.

3.2.3. Pre-screening primers to optimise PCR conditions

Each primer pair was initially tested on genomic pig and hamster DNA and a negative control of double distilled water, to ensure that the amplification was

specific to porcine DNA only and also to optimise the PCR conditions and eliminate non-specific DNA binding.

It became apparent that some primers worked well in the pre-screen, however when the same PCR conditions were used over the radiation hybrid panel, the results were much weaker. It was therefore decided that a DNA sample, which contained 20 % pig DNA and 80 % hamster DNA would also be included in the pre-screen test. This was shown to represent the rough proportions of porcine DNA in the panel at the lowest level and to give a clearer indication as to whether the conditions optimised in the pre-screen were likely to be successful when the panel was screened.

The control pig (RH09conpig) and hamster (RH09conA23) DNA samples (ResGen™ Invitrogen corporation, Paisley, UK) were diluted from the stock concentration of 25 ng μl^{-1} to 3 ng μl^{-1}, the same concentration as the DNA in the panel. In addition, a mix of 20 % pig DNA and 80 % hamster DNA was also made up to 3 ng μl^{-1}.

For each pre-screen, a single primer pair was tested over pig DNA, hamster DNA, 20:80 pig/hamster DNA and double distilled water. Target DNA fragments were amplified in 20 µl reaction volumes using 30 ng of diluted DNA samples (3 ng μl^{-1}). The standard reaction mix included 5 pmol of each primer (MWG Biotech, Germany), 2.0 mM of each of dATP, dTTP, dGTP and dCTP (Amersham Pharmacia Biotech inc, Little Chalfont, UK), 1.5 mM $MgCl_2$ in 1x PCR buffer (Roche Diagnostics, Mannheim, Germany), 5.7 % sucrose and 0.845 µg creosol red indicator dye (Aldrich Chem Co., WI, USA) and 0.45 U *Taq* DNA polymerase (Roche Diagnostics, Mannheim, Germany).

Where additional magnesium was necessary in the reaction mix, the required amount of double distilled water was substituted with $MgCl_2$ (25 mM) (Boehringer, Mannheim, Germany). In order to ensure consistency in the preparation of the PCR reaction mix, bulk master mixes were made and frozen at -20° C. The master mixes

were made at three different magnesium concentrations (1.5 mM, 2.0 mM and 2.5 mM) and contained all the ingredients of the PCR reaction mix apart from the primers and the *Taq* enzyme (Table 3-1). Note that the same master mixes were used when screening the primers over the radiation hybrid panel.

Table 3-1 PCR master mixes at three different magnesium concentrations

PCR master mix ingredients	**Volumes required for 1.5 mM mix**	**Volumes required for 2.0 mM mix**	**Volumes required for 2.5 mM mix**
10x PCR buffer	4.4 ml	4.4 ml	4.4 ml
34.6 % w/v sucrose	7.26 ml	7.26 ml	7.26 ml
Creosol red	2.2 ml	2.2 ml	2.2 ml
dNTPs	4.4 ml	4.4 ml	4.4 ml
Double distilled water	2.44 ml	1.56 ml	0.68 ml
25 mM $MgCl_2$	N/A	0.88 ml	1.76 ml

From the master mix, 1035 µl of each mix was aliquoted into 20 individual 1.5 ml eppendorf tubes. Each tube of master mix contained enough reagent to screen one 96-well panel DNA plate or to pre-screen at least 20 primer pairs.

The PCR reactions were carried out on a Peltier thermal cycler PTC-225 DNA engine tetrad (MJ Research), using the same PCR program for every primer tested. After an initial 94° C denaturation step for 5 minutes, the anneal phase consisted of 36 cycles of 94° C for 30 seconds, the primer specific anneal temperature for 50 seconds and 72° C for 50 seconds and finally an extension cycle of 72° C for 5 minutes.

The PCR products were separated by gel electrophoresis through 400 ml 2 % Ultra pure DNA grade agarose (Bio-Rad laboratories, Hemel Hempstead, UK) in 1x TBE buffer, stained with 160 µg ethidium bromide (Sigma-Aldrich, Steinheim Germany). The products were run with either a PhiX174/*Hin*fI marker or PCR marker G3161 (Promega corporation, Madison WI) to check that PCR products of the expected size had been amplified. The gel was visualised on an ultra-violet transilluminator and photographed.

3.2.4. *Preparation of radiation hybrid DNA plates*

The DNA for this specific radiation hybrid panel is available from Research Genetics (ResGen™ Invitrogen corporation, Paisley, UK). This panel consisted of 94 individual hybrid clones at a concentration of 25 ng μl^{-1}. The hybrid DNA (900 µl of each) was stored in 1.2 ml micro-tube cluster plates (ABgene, Epsom Surrey, UK) at -80° C. The remaining two wells each contained the same pig and hamster control DNA as used for the pre-screen test. The plate format of each hybrid clone is shown in Figure 3-3.

	1	2	3	4	5	6	7	8	9	10	11	12
A	D6	E8	FF1	HH7	J6	L6	N6	O6	Q5	S1	S12	P
B	D1	E4	F12	HH1	J5	KK8	N5	O5	Q2	R12	S11	H
C	C12	E3	F6	H11	I12	KK3	N4	O2	Q1	R5	S10	Z1
D	BB2	E1	F4	H10	I11	KK2	N3	N12	P11	R1	S8	Y8
E	B9	DD5	F1	H4	I5	K8	N1	N11	P6	Q12	S5	W5
F	B3	D12	EE4	H3	I4	K6	M10	N10	P4	Q9	S4	T12
G	AA5	D10	E11	FF7	I2	K2	LL4	N8	O9	Q8	S3	T11
H	AA1	D9	E9	FF4	HH8	J11	LL1	N7	O8	Q6	S2	T5

Figure 3-3 96-well plate format of radiation hybrid panel DNA clone identifiers. (P – Pig genomic DNA, H –Hamster genomic DNA).

The DNA in every well of one stock DNA plate was diluted to 3 ng μl^{-1} with dilution buffer into twelve 96 deep-well plates (ABgene, Epsom Surrey, UK) using a Hydra™ 96 dispenser (Robbins Scientific Corporation, Sunnyvalle, CA). The same dispenser was used to aliquot 10 µl of each copy of the diluted panel into 96-well skirted plates (ABgene, Epsom Surrey, UK) and these were sealed with microplate sealers (Greiner Labortechnik Ltd, Stonehouse, UK) and stored at –20° C.

3.2.5. Screening primers over radiation hybrid panel

The PCR master mix was made as described in section 3.2.3 and each primer pair was tested in duplicate over the whole panel. To each tube of master mix (total volume 1100 µl), 5 pmol of each primer and 0.45 U *Taq* DNA polymerase were added and vortex mixed. Using a Multipette® plus dispenser (Eppendorf, Hamburg Germany) 10 µl of the PCR mix was dispensed into every well of two plates containing the DNA panel (3 ng μl^{-1}). The PCR was run with the same conditions as outlined in section 3.2.3 and the resulting PCR products from both plates run together on a 400 ml 2 % agarose gel and visualised on an ultra-violet transilluminator.

Each hybrid in the panel was scored for the presence (+) or absence (-) of an amplified porcine PCR product. Any primers, which amplified extra bands, were disregarded. Primers were only considered successful, where there were no more than three discrepancies between the duplicate screen plates. These differences between the two plates were scored as a question mark. It is important to score the RH panel in duplicate because unlike with meiotic mapping there are no automatic checks to ensure that the bands amplified for each marker represent the two alleles inherited by the F1 individual from each of the two parents. Therefore performing the PCR twice and comparing the results will help ensure that the scoring of the panel was accurate. Clearly the well containing the porcine genomic DNA should always amplify a band of the specific size for the marker being typed (the same size band will be present in each of the positive clones). In contrast the control hamster DNA should not amplify a PCR product of a similar size to that of the porcine PCR product.

All the results for markers, which were successfully typed, were recorded in the Roslin radiation hybrid database (http://www.ri.bbsrc.ac.uk/radhyb) for subsequent analyses.

3.2.6. *Statistical analysis*

3.2.6.1. Use of CarthaGene to construct radiation hybrid maps

CarthaGene version 0.5 (http://www.inra.fr/bia/T/CarthaGene/) was the software chosen to construct the radiation hybrid map of SSC8. The diploid radiated hybrid data produced was used and the order of the markers and the distances between them determined.

Some mapping packages only investigate two point measures, that is they search locally for linkage between just one pair of markers at a time. However with these programs there are problems with the two point estimations when two groups of markers are far apart from one another. In contrast CarthaGene combines these local search techniques with multipoint maximum likelihood criteria that defines the best marker order using the data from all markers simultaneously (Schiex and Gaspin, 1997).

The successful marker scores over the 94 hybrid clones of the panel were converted from the stored format in the radiation hybrid database to a CarthaGene compatible input "dat" file (i.e. positive clones were labelled "H", negative clones "A" and discrepancies marked with a "-"). The data set of successfully screened markers was then loaded into CarthaGene.

3.2.6.2. Identification of linkage groups using pair wise analysis

The command "mrklod2p" was used to obtain a two-point LOD score matrix for each marker against all other markers. This analysis highlights where a pair of markers is completely linked. If linked markers were included in the CarthaGene analysis, the computational time would be significantly lengthened because the algorithms used would not be able to determine the order of the two markers relative to one another. One of the pairs of the linked markers was therefore removed from the subsequent map building process and then added back to the final map after it had been built, at the same position as its linked marker.

All the markers were then sorted into their linkage groups using a LOD threshold of 4.0 and by searching a local two-point threshold distance of 0.2 rays. Any single marker, which did not belong to a linkage group, was then excluded from the map building process. Either these single markers did not map with the markers being typed on the chromosome of interest or they mapped too far from established linkage groups and consequently the LOD score was too low to be certain of the position of the marker relative to other markers on the map.

3.2.6.3. Build framework and comprehensive maps of SSC8

A pre-defined set of algorithms within the "defalgo" program was then run over the data set for each separate linkage group and the order of the markers and the distance between them determined. The specific algoithms used are shown below:

nicemapl
annealing 100 100 0.1 0.9
flips 5 3.0 1
flips 7 3.0 1
polish

The first stage of the "defalgo" program builds a framework map from the raw data set using a "heuristic" algorithm called "nicemapl". Using the two points LOD criteria only, the algorithm calculates a rough order and distances between markers. Starting from the best framework map built, the local search algorithm "simulated annealing" then uses the multipoint maximum likelihood criterion to build more accurate maps.

The specifications of the annealing algorithm determine the accuracy of the map whilst keeping the computational time to a reasonable level. In the program used for this analysis the following specifications were used; 100 moves, an initial temperature of 100°, a final temperature of 0.1° and a cooling factor of 0.9. The theory of using decreasing temperatures is that it mimics crystallisation as a heated solid cools and solidifies. In the early stages the material has more "choices" about its structure and does not follow the "optimal" expected energy path. At the starting

temperature of "100° C", there were a set number of map solutions and at each step of decreasing the temperature new map solutions were created, including changing the order of distant markers. The annealing process built a map at each temperature and printed a "+" where the calculation found a better map solution. At the final temperature of "0.1° C" there was no probability of finding a non-optimal map solution.

The best map found was then displayed along with its log likelihood. This map was tested for accuracy using a "verification algorithm". This involves flipping pairs of markers within a specified window size. The number of markers within the window was specified. Also specified was a threshold by which the new map is said to be more accurate; this is the minimum difference in log likelihood between the new map built and that of the best map prior to flipping. The flips analysis was then repeated with a higher window size to ensure that more distant markers were correctly placed relative to one another.

Finally the "polish" command was used. This takes each locus in turn and tests every possible position of this locus on the map, whilst retaining the order of the other markers, to test whether alternative positions improve the overall accuracy of the map. The maps produced throughout the analysis were then displayed in order of their log likelihood values, the map with the highest log likelihood being the most accurate. It was then possible to display the details of the most accurate map produced after the flipping and polishing stages of the analysis. This output includes the marker order, distances between markers in centiRays and the two point LOD scores between adjacent marker pairs.

3.2.6.4. Calculating marker retention frequencies

The data set for all successfully mapped markers was retrieved from the Roslin radiation hybrid database (http://www.ri.bbsrc.ac.uk/radhyb) and analysed using RHMAP two-point analysis (Boehnke et al., 1991) (available from the Roslin radiation hybrid website). Part of the analysis gives a table of the haploid marker retention frequencies for each marker along the chromosome. This is the percentage

of positive hybrids from the total number of hybrids successfully typed (there may be up to three discrepancies). It is assumed when building radiation hybrid maps that the marker retention frequency is random across the chromosome i.e. that the breakage along the chromosome is random (Lunetta et al., 1995).

RHMAP gives the output of marker retention frequency for haploid data. This therefore needs to be corrected for this data being diploid. The following calculation was used to convert the haploid marker retention frequencies to diploid values:

$$p_D = 1-q_D, \text{ where } q_D = \sqrt{q_H} \text{ and } q_H = 1-p_H$$

p_H = haploid retention frequency, p_D = diploid retention frequency

3.3. Results

3.3.1. Identification of DNA markers on SSC8

There were three classes of markers from pig chromosome 8 that were typed over the radiation hybrid (RH) panel. Firstly 50 primer pairs were obtained for anonymous DNA microsatellite sequences previously identified to map to porcine chromosome 8 and for one microsatellite within the *SPP1* gene. The published references for these primer sequences were found from either the Roslin pig genome database (Hu et al., 2001) (http://www.thearkdb.org/browser?species=pig) or the USDA-MARC swine genome mapping database (http://sol.marc.usda.gov). These microsatellite markers covered the whole of chromosome 8 in order to establish a framework map, on which to map genes to the chromosomal regions of interest. The microsatellite marker within the *SPP1* gene was the same primer pair as used to map this gene to the linkage map (Table 2-2).

Secondly 14 primer pairs were designed for porcine genes already shown to map to SSC8, either from the linkage map produced in chapter 2 (*IBSP, GNRHR, AREG, FGG, HD, QDPR, SLIT2* and *STE* on Figure 2-4) or from other published data (*EGF* (Spotter et al., 2002), *SPARCL1* (Maak et al., 2001), *MAN2B2* (Okamura et al., 1995), *CPE* (Cargill et al., 1998), *PDGFRA* (Moller et al., 1996) and *KIT* (Sakurai et al., 1996)). The same primer pairs for the genes *AREG, FGG, HD, QDPR* and *SLIT2* were used for RH mapping as for the linkage map construction (Table 2-2). *IBSP, GNRHR* and *STE* were also present on the linkage map, however alternative primers used to map *IBSP* and *GNRHR* were successfully designed from non-coding regions of porcine sequences from the comparative mapping search of human chromosome 4 (see Table 4-1). Published primer sequences for a microsatellite repeat within the porcine estrogen sulfotransferase (*STE*) gene were found when searching the NCBI website (http://www.ncbi.nlm.nih.gov/) for nucleotide sequences of genes already known to map to SSC8.

Finally primers were designed within 24 porcine sequences that showed homology to human 4 genes (the identity of these genes and the methodology of marker

development are detailed in section 4.3.3). When designing these primers for radiation hybrid mapping, the amplified region of the gene needed to be ideally within intronic genomic sequence or within the 5' or 3' untranslated regions (UTR). This was done to avoid the amplification of homologous coding regions from the hamster DNA present in the panel. The exonic regions of genes are highly conserved across species, whereas the non-coding regions are much more variable.

3.3.2. Marker pre-screening

The primers for all 88 markers outlined in section 3.3.1 were pre-screened with control pig and hamster DNA, the 20 % pig: 80 % hamster DNA mix and water. PCR was carried out at various temperatures and with 1.5, 2.0 and 2.5 mM magnesium in the PCR mix in order to optimise the conditions required. Where a single band was amplified with the porcine DNA and not the hamster or water, the PCR was considered successful. In order for the subsequent screen of the RH panel to work, the band amplified with the 20 % pig DNA mix was required to be sufficiently bright. If the band was faint, then the PCR was repeated at a lower annealing temperature. Where non-specific bands were amplified, either the temperature was increased or the magnesium concentration decreased to give a single bright band of the correct size for the marker amplified.

Figure 3-4 shows a couple of examples of pre-screen gels, where the magnesium concentration was adjusted to optimise the PCR conditions. The marker for *BMPR1B* was successfully screened at 2.5 mM magnesium and an anneal temperature of 58° C and the marker for *UNC5C* was successfully screened at 1.5 mM magnesium at 60° C.

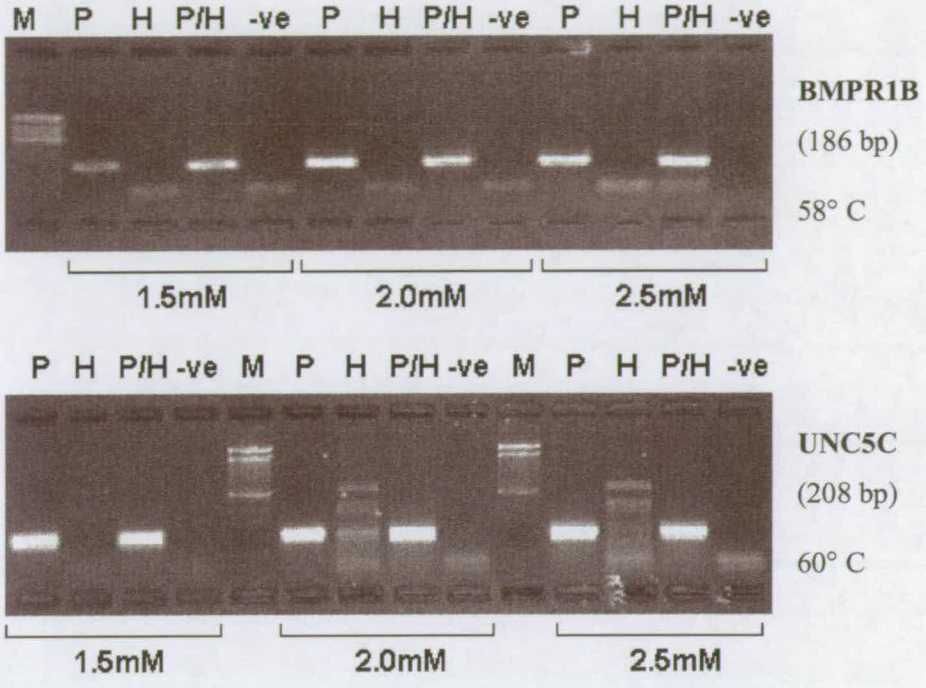


Figure 3-4 Pre-screen gels for markers *BMPR1B* and *UNC5C*. (P =pig DNA, H =hamster DNA, P/H =20 % pig: 80 % hamster DNA mix, -ve =negative control (distilled water) and M =DNA size marker).

3.3.3. Screening of markers

A total of 70 out of the 88 primer pairs successfully passed the pre-screen test. The optimised PCR conditions for each marker were then used to screen the radiation hybrid panel (94 hybrid clones, control pig and hamster DNA). In order to eliminate markers that had been retained more often than would be expected for a single chromosome within the panel, I decided to exclude markers with a retention frequency greater than 50 % (47 positive bands). Also any markers with a retention frequency that was too low, i.e. less than 3.2 % (4 positive bands) were also excluded. Some markers which appeared to have amplified bright bands in the pre-screen, had bands in the RH panel which were too faint to accurately type. Following the procedures established by groups who have developed RH maps of human and other species, only those markers that showed three or less discrepancies between the two duplicate screens of the panel were included, therefore several markers were also eliminated on this condition.

A total of 56 markers were successfully typed over the panel. This included thirty-six microsatellites, eight genes previously known to map to SSC8 and twelve porcine sequences from homologous human chromosome 4 genes. Figure 3-5 is an example of duplicate screen gels for the microsatellite marker *SW905*. Each well represents the 94 hybrid clones of the radiation hybrid panel, the well labelled "H" contains the product from the PCR on hamster DNA and the well labelled "P" contains the product from the PCR on porcine DNA. An amplified band shows that the clone is positive for the marker being typed. The pattern of marker retention between clones is used to determine the distance i.e. the amount of breakage between each marker on the chromosome typed. Table 3-2 shows the primer sequences, PCR conditions and references to where the porcine sequences were obtained or designed for all 56 successfully typed markers.

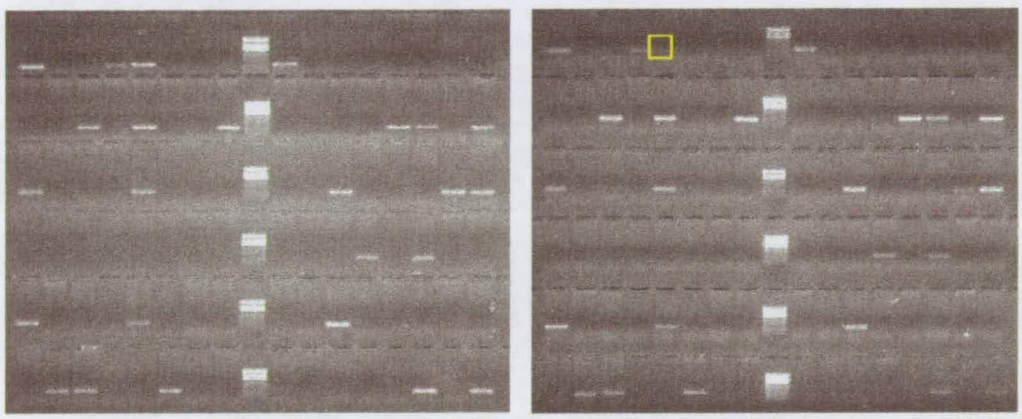


Figure 3-5 Example of duplicate screen gels for the microsatellite marker *SW905*. There was 1 discrepancy between gels (highlighted with yellow box). (P =pig DNA and H =hamster DNA, the middle lanes contain a DNA size marker).

Table 3-2 Primer sequences, PCR conditions and references to where the porcine sequences were obtained for all 56 markers that were successfully typed over the radiation hybrid panel. The markers are grouped according to where they were obtained. (See section 3.3.1 for more detail).

Marker	F primer	R primer	Anneal temp.	Mg conc. (mM)	Reference to the source of the marker or gene sequence	Name of porcine gene
Microsatellites						
SPP1	GCTAGTTAATGACATTGTACATAA	GTGTCATGAGGTTTGTGCCACTGC	58	1.5	(Moran, 1993)	Secreted phosphoprotein 1
S0086	GCACAGTCTATTGATACTGGCGTC	CTGAGAACTTCCATATGCTCCTGG	58	1.5	(Ellegren et al., 1993)	
S0144	ATTAGAGCCACCTGAAGGCT	CGTTGGTCGGCAGCTATAGC	55	1.5	(Jung et al., 1994)	
S0225	GCTAATGCCAGAGAAATGCAGA	CAGGTGGAAAGAATGGAATGAA	55	1.5	(Robic et al., 1994)	
S0285	TAGCCTGCCCTCAAAAGACCT	TGCCCTAAAAATCCCCTGTTC	58	2.0	(Rohrer et al., 1996)	
S0376	TCTGGGCCACTGAAGGACTT	TGACTGCAGCTCCGATTCAA	58	1.5	(Robic et al., 1997)	
SW1037	ACATTTCATGGGAAAAATAAGG	ACACAGCCCTACTGACATTTTG	50	1.5	(Alexander et al., 1996a)	
SW1312	TTGGTGACAAAGAGGCAAT	GAGACATGCAGTTTAGCTGCC	58	2.0	(Alexander et al., 1996a)	
SW1649	CACAGGTCCATCACCTCATG	TGGGTATCTAATGTACATCACGG	58	1.5	(Alexander et al., 1996a)	
SW1679	GCCAAGAGGGAAGCTTATAGC	CAAATCCATTTAGATGGTAGGC	58	1.5	(Alexander et al., 1996a)	
SW1702	ACCTCCATATACTGCAAGCG	TACCAGTCGAGTTCGCTGAC	58	1.5	(Alexander et al., 1996a)	
SW1980	GCTTCTGTATGCCACAGCTG	CCCCCATTTGAACAATGAAG	58	1.5	(Alexander et al., 1996a)	
SW368	ACATTAAATAGGATTATATATGCATGGG	TTCATTTTTTTTAAAGTCAGTTATCTCA	55	2.5	(Alexander et al., 1996a)	
SWR1921	TGCTCGGGTTTGTTAACCTC	CCTTCTATAGCCTTTCCAATTTATG	58	2.0	(Alexander et al., 1996a)	
SW2611	CTTGTTTCCCGCAGTCTCTC	GTGTGTTCCAGATGAACCTGG	60	2.0	(Alexander et al., 1996b)	
SW1070	CTTGCAGCATCACTCTTAGGC	TCTATGTGCCTTGGAGTGAGG	58	1.5	(Rohrer et al., 1994)	
SW1080	GGGAAATTTGGATTTGAAATTG	TCCCTGTCACTGTAACTTGCC	53	1.5	(Rohrer et al., 1994)	
SW149	TCATGTTCACAGAACACCTTCC	AGTTACGGTGGGTCGTAATCC	55	1.5	(Rohrer et al., 1994)	
SW171	CAGATTGTTTAGCCTCTGTGTG	CATCTTTTCCAATGACAACATG	58	1.5	(Rohrer et al., 1994)	

SW194	TGCCTGCTTACATTAAGTGGG	CCTCACTTAAGAAGGTTCCTGC	62	1.5	(Rohrer et al., 1994)	
SW205	CACAGGTCCATCACCTCATG	GGGTATCTAATGTACATCACGG	58	1.5	(Rohrer et al., 1994)	
SW211	TCATCAAGAAAATTGGCTTGG	TGACCACAAGGAAGAAACTGG	60	1.5	(Rohrer et al., 1994)	
SW268	CTGATTCACTTTCATTCGAGAA	AGCCCTTCCCTTAATATAACCC	55	1.5	(Rohrer et al., 1994)	
SW29	AGGGTGGCTAAAAAAGAAAAGG	ATCAAATCCTTACCTCTGCAGC	60	1.5	(Rohrer et al., 1994)	
SW374	AGTAATCCCATCCTCCCCAG	TGCTCTCCAGCCCTCAAG	62	1.5	(Rohrer et al., 1994)	
SW444	ATAGTTTCGGTTGGCCCAG	CTTAAGCCTCAAGCTAACAGGC	58	1.5	(Rohrer et al., 1994)	
SW61	GAGAGGGATGAGCACTCTGG	AGAGCATTCCAGGCTTCTCA	62	1.5	(Rohrer et al., 1994)	
SW7	TAACCATGCTTTTCCTAGGTGG	CCAGAGCTGAGTAAAAAGGTCA	60	1.5	(Rohrer et al., 1994)	
SW763	GGGTGCATTGTTCTCATATGG	TGCTCTAGCAACACACACCC	62	1.5	(Rohrer et al., 1994)	
SW905	ATCCCAACCTTCTTTCAAAGG	TCCAGTGGCAGAACAACATG	60	2.5	(Rohrer et al., 1994)	
SW1345	CCTGTGCCCAGTTCCATC	CATTGACTCCAGGTAGAGTCCC	58	1.5	(Rohrer et al., 1996)	
SW933	ACATATACTTCCGACAGCCCC	AAGAGCTTGGTGAATTGAGAGC	57	1.5	(Rohrer et al., 1994)	
SW1924	GTCAAACTTTGCAAATGTTATGTG	TTGACAAAACTAGGTAAAGCATGG	60	1.5	(Rohrer et al., 1996)	
SWR750	CATGGACATTAAAAAAAGTGGTC	GGAACCTCCATGTGCCTG	58	1.5	(Rohrer et al., 1994)	
SW1085	CAGGCTCCCTGACTTCAGAC	TAGGTCCATCCATGTTTCTGC	58	1.5	(Rohrer et al., 1994)	
S0069	TGCAAAACTAATGTTTGTGTTGCC	CATATGCCACAGGTGTGACCTAAA	60	1.5	(Fredholm et al., 1993)	

Porcine sequences from genes in region of human chromosome 4 homologous To SSC8qarm (between *GNRHR* and *CPE* on HSA4) (See Figure 4-2)					Porcine accession number (region of primer design within gene)	
AMBN	GGTCCTCTTCTTGCTTTCATTC	GCAGTGTCACATTTCCTGGAT	60	1.5	U43404 (3'UTR)	Amelioblastin
CSN1	GGAAATCGCAAATCGAAGTC	TGGCACTTACAGGAGAAGCA	58	2.0	X54973(3'UTR)	Alpha casein
CSN10	CAAACTCTCCTTTGGCCAGT	CATTTGCATTCAATTGGCTTT	58	1.5	X51977 (3'UTR)	Kappa casein
IL8	GCAGTGCTGTATTGAATGACG	TCTCTTTGGAAATACATAAAAACAAA	52	2.0	M86923 (3'UTR)	Interleukin 8
SCYB6	TGCACAGTTCTGGCTAAGGTT	CAAAGCACAGAAGTACTATCAACCA	58	1.5	M99368 (3'UTR)	Alveolar macrophage-derived chemotactic factor II (AMCF-II)
IL2	CCCACCCCCTTAAAGAAAGA	TACCTGTGTGGCAAAAAGCA	56	2.0	AB041341 (5' UTR)	Interleukin 2
BMPR1B	GGGCAAGAACAAGCATCTCT	GTTTCTTCCTCCTGCGAACA	58	1.5	Obtained primer sequences from Gary Rohrer (intronic genomic)	Bone morphogenetic protein type 1 receptor
UNC5C	AGACAGTGCGCAGGTTAGGT	CACGTTCAGCCACAAACAAG	60	1.5	BI182276 (3' UTR)	PIG EST UNL-P-FN-bh-a-07-0-UNL.s1
DSPP	CAATATCAGAGCCCCTTGCT	TCATGCTACCACCAGGGAAG	59	2.0	AF332578 (3' UTR)	Dentin sialoprotein
COPS4	GTCATGATGCAAATGCCTTG	AACGTGGGACAAACAGATCC	59	1.5	BF702060 (3' UTR)	PIG EST MI-P-A1-aao-d-06-1-um.s1
PMBP	GGCTTGCCTAGAAACCGAAA	CCTGGCTGGTAATGCGTACT	58	1.5	X99714 (3' UTR)	Steroid membrane binding protein
NEK1	AGATGCCAAAGAGCAGGAGA	AGTGGCAGAGAAGGACGTGT	54	2.0	X91312 (3' UTR)	PIG EST clone K27 from ovary

Genes already known to map to SSC8						
GNRHR	GCAAACACCTGATGTCTAGAATGT	AATGCACCAGATTCCCAAAG	60	1.5	AF227686 (intronic genomic)	Gonadotrophin releasing hormone receptor
IBSP	CCAACGTGGTCTTCTGGACT	TGTTTTAGCAGTGGGTGCAA	55	1.5	L10363 (3'UTR)	Bone sialoprotein
SPARCL1	GCGAGCAGGAAGAAGACA	TTGGGCTTAGTGATTTTTGAG	60	1.5	(Maak et al., 2001)	SPARC-like 1 (mast9, hevin)
SLIT2	ACTACCAAAAGCAGCAGGGCTATG	TCAAATATTTCCACTATGAAGCATTCA	60	2.5	PCR- RFLP-on linkage map	SLIT Drosophila homolog 2
STE	GTCTTTCCATTTATTTCAGTTGATTTG	TCTGCATATTTTTCTGTAGTGATACATC	58	2.5	AF406989 (published microsatellite primers)	Estrogen sulfotransferase
MAN2B2	GAAGGGCAGTGAGCAAGTTC	ACCCCTTGGGGTAAGAAGTC	56	2.0	D28521 (5'UTR)	Alpha mannosidase
CPE	CTGCACCTGGAGCATATGAA	GTACCTGTACCGGGGGTTTC	58	1.5	AF021874 (intronic genomic)	Carboxypeptidase E
EGF	GAAACAATTCCCGTGTTCTCT	TCACTTCCACACCTGTAACATCT	58	2.0	(Spotter et al., 2002)	Epidermal growth factor

3.3.4. Use of CarthaGene to construct radiation hybrid map of SSC8

3.3.4.1. Linkage groups and LOD thresholds

The typing for the 56 markers were loaded into CarthaGene and a two point LOD matrix produced. This analysis highlighted that three pairs of markers were completely linked, *IBSP* and *DSPP*, *IL8* and *SCYB6* and *SW1702* and *SLIT2*. Markers *IBSP*, *IL8* and *SW1702* were therefore removed from the map building process to speed up the computational time. The markers were then sorted into six linkage groups, three of which contained single markers. These were *S0069*, *NEK1* and *PMBP* and these markers were removed from subsequent analyses.

The porcine sequence obtained for steroid membrane binding protein (nucleotide accession number X99714) had shown a low level of sequence identity to the human sequence for *PMBP* (accession number XM_003419). The alignment score was 64 bits and in order to be reasonably confident of a match between sequences the score ideally needs to be greater than 80 (for more detail see section 4.2.1). Therefore the porcine sequence was then blast searched against human GenBank, EMBL, DDBJ and PDB databases and the closest match to the sequence was to the gene *PGRMC1* (progesterone receptor membrane component 1) that maps to human chromosome X. In fact the entry for human *PMBP* (XM_003419) has since been withdrawn from the database. It was realised that *PGRMC2* (accession number NM_006320) actually maps to human chromosome 4q28.2.

Therefore the human sequence for *PGRMC2* was BLAST searched against the GenBank, EMBL, DDBJ, PDB and EST database for homologous porcine sequences. An EST sequence (accession number BI400947) with an alignment score of 301 BITS was found and primers were designed within the 3' end of the sequence. Although the PCR conditions for this primer were optimised, there were too many differences between the duplicate plates (i.e. more than three discrepancies) to be able to include it in subsequent analyses.

The porcine marker for the gene *NEK1* had shown a high level of sequence identity to the human gene, but did not fit within the linkage groups on SSC8. It is possible that the match was to a porcine EST sequence for a gene that does not map to SSC8 or that the gene maps too far from other markers on the map to confidently place it within the LOD threshold of 4.0. Figure 3-9 shows that the microsatellite marker *S0069* maps to the linkage map relatively far from the markers *SW1679*, which maps to the top RH linkage group and *SW171*, which maps to the middle RH linkage group. This would explain why it was not placed within a group of markers using a LOD threshold of 4.0 and a search distance of 0.2 rays.

3.3.4.2. Draft map of 53 markers on SSC8

The remaining three radiation hybrid linkage groups of markers were then used to construct the RH map of SSC8. It was not possible to map each group relative to one another; therefore three separate maps were built for the markers in each group. The markers within each linkage group, their position on each map in centiRays and the two point LOD score of adjacent markers is shown in Figure 3-6. Where the LOD score between a pair of markers is lower than the threshold of 4.0, the confidence of placing those markers adjacent to one another is reduced. This usually occurs where the distance between markers is large, for example the distance between markers *SW268* and *SW933* in linkage group one is 0.9 rays.

For each linkage group, the order of markers for the maps built throughout the CarthaGene analysis are shown in order of log likelihood in Figure 3-7. The top map with the highest log likelihood value is the most likely order of the markers determined after the flips and polish analysis.

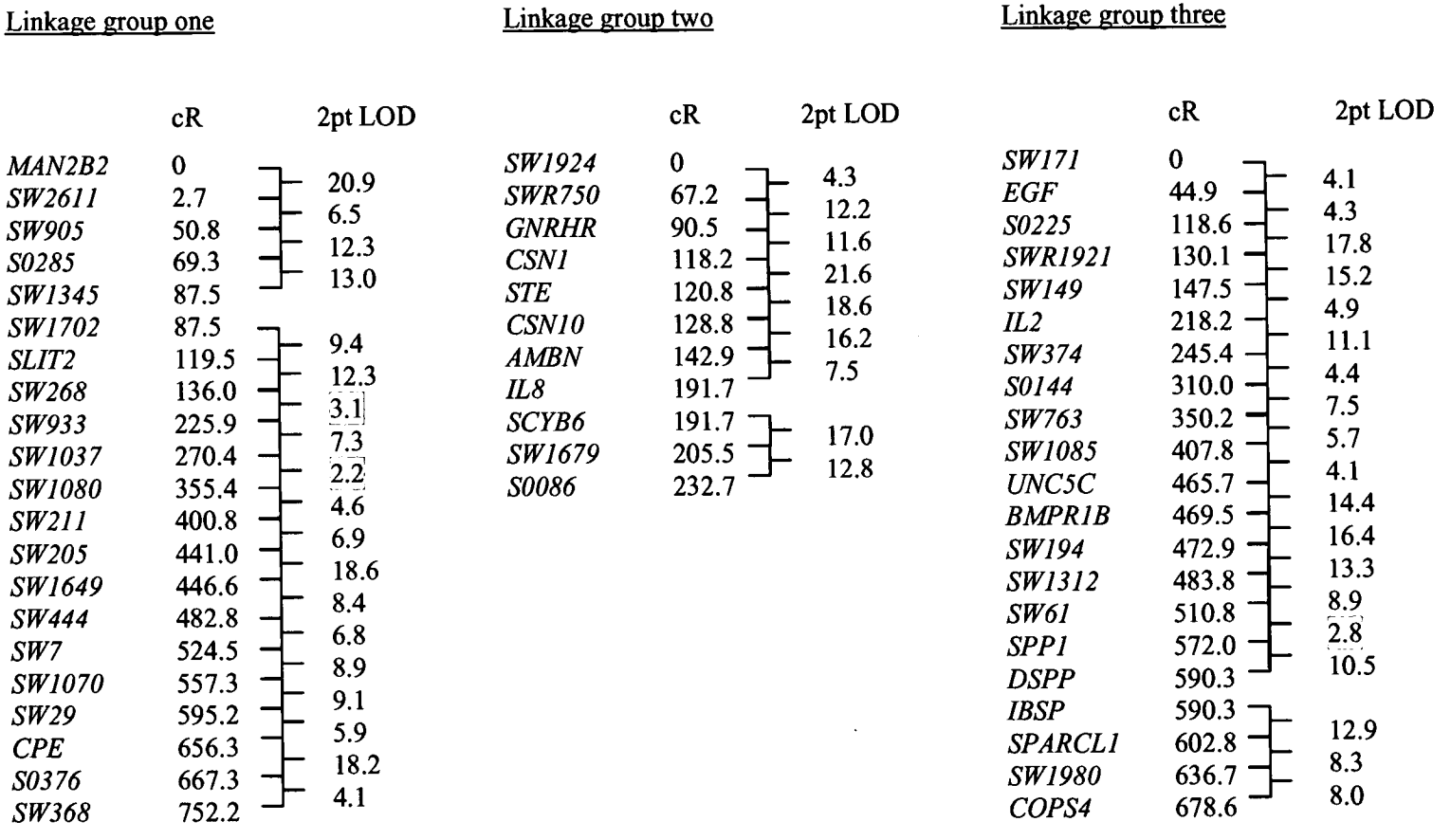
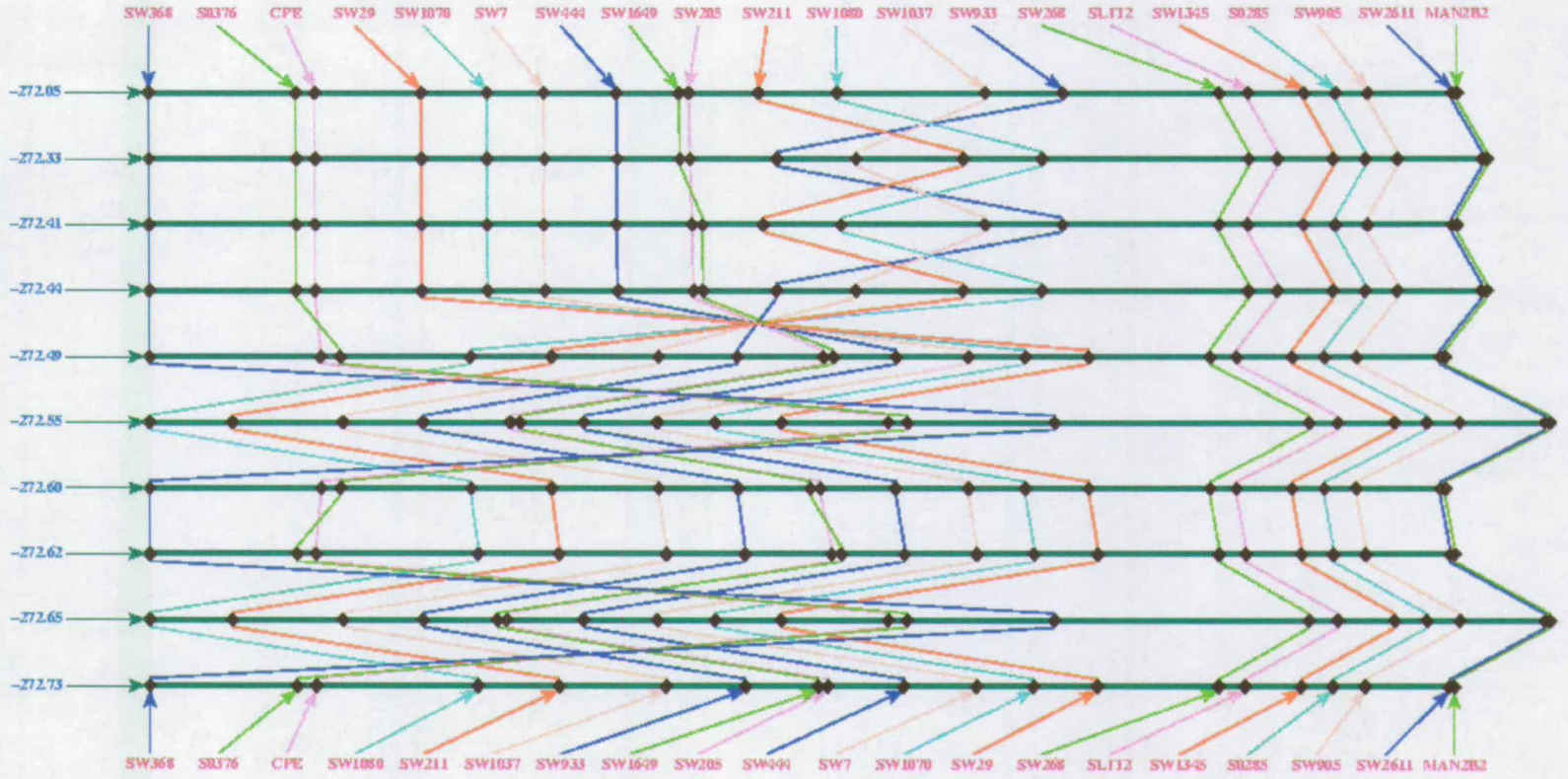
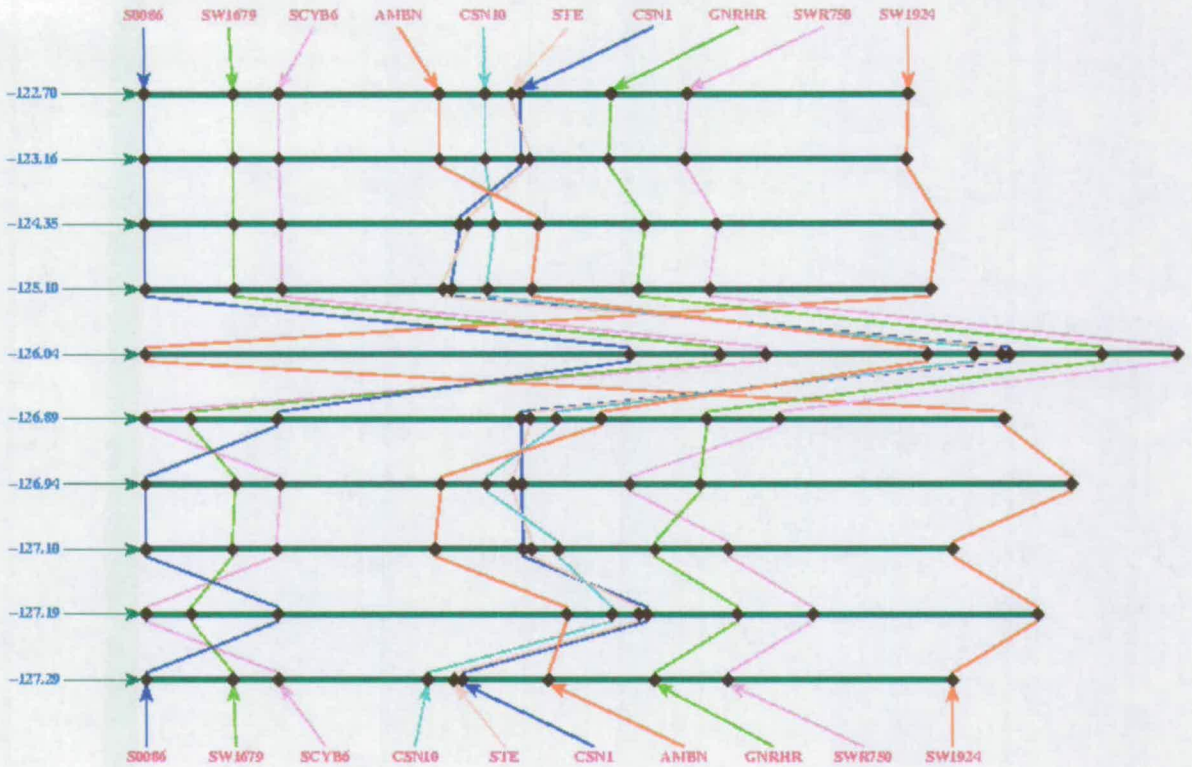


Figure 3-6 Maps of the three linkage groups on SSC8, showing the position in centiRays of each marker and the two point LOD scores between adjacent markers. LOD scores lower than the threshold of 4.0, between neighbouring markers are highlighted in yellow.

Linkage group one



Linkage group two



Linkage group three

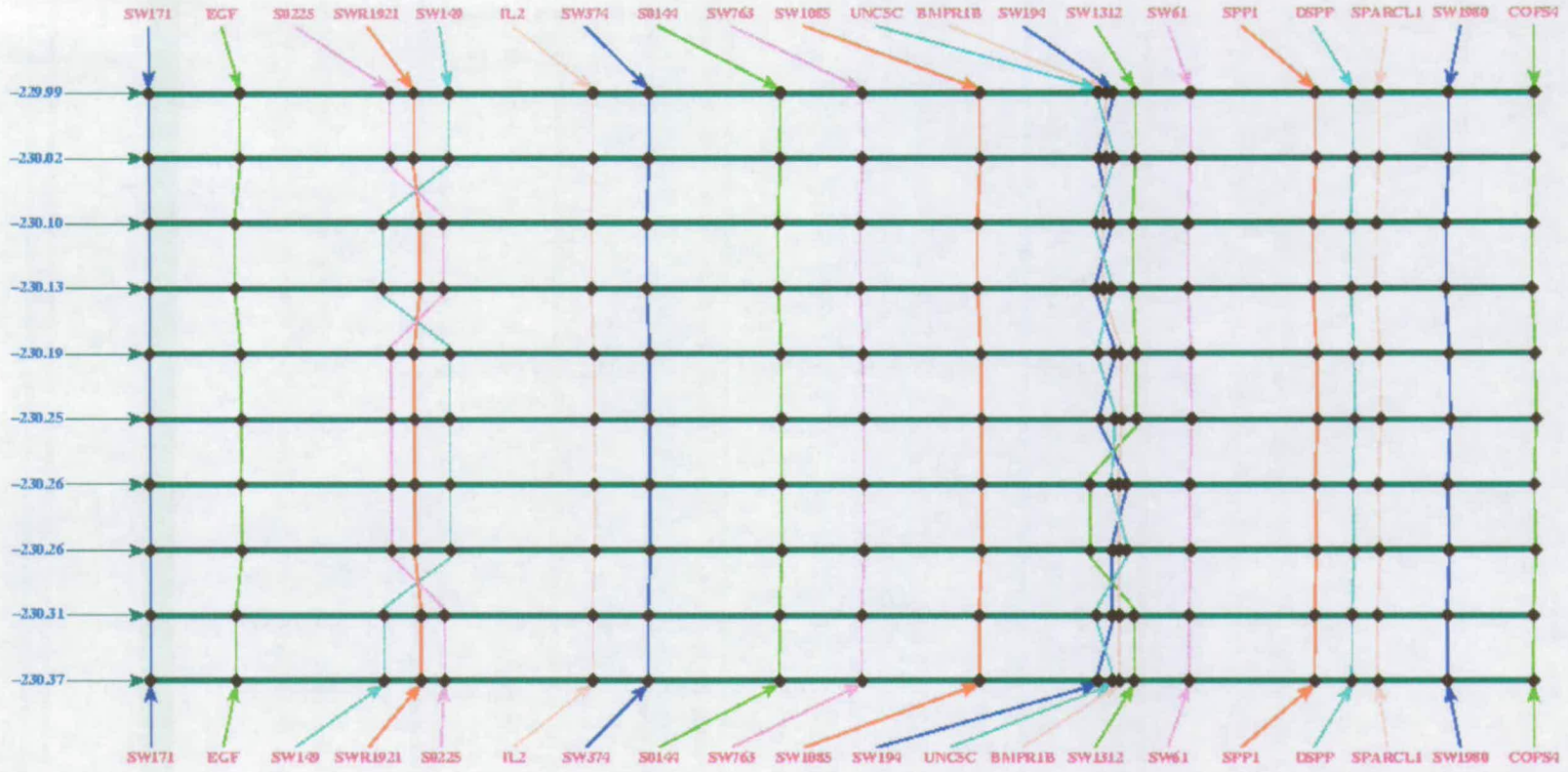


Figure 3-7 Log likelihood plots of marker order for the three linkage groups on the radiation hybrid map of SSC8.

By aligning these three most accurate maps with the linkage map built for the QTL analysis (Figure 2-4), it was possible to position the three groups relative to one another (Figure 3-8). This alignment was done using the Anubis map viewer (http://www.thearkdb.org/). The distance between the linkage groups was unknown and was therefore arbitrarily set to be 100 cRays. This final map consisted of thirty-five microsatellite markers, eight markers for porcine genes known to map to SSC8 and ten genes predicted to map to SSC8 from comparative mapping information of genes on human chromosome 4 and the total map length was around 1900 cR3000.

Figure 3-9 compares the radiation hybrid map built to the linkage map of porcine chromosome 8 produced by the USDA-Meat Animal Research Center (http://sol.marc.usda.gov/). Many of the same microsatellite markers are present on both maps and it is therefore possible to not only compare the level of marker order similarity, but also to compare how physical distances between markers on the radiation hybrid map compare to genetic or recombination distances on the linkage map.

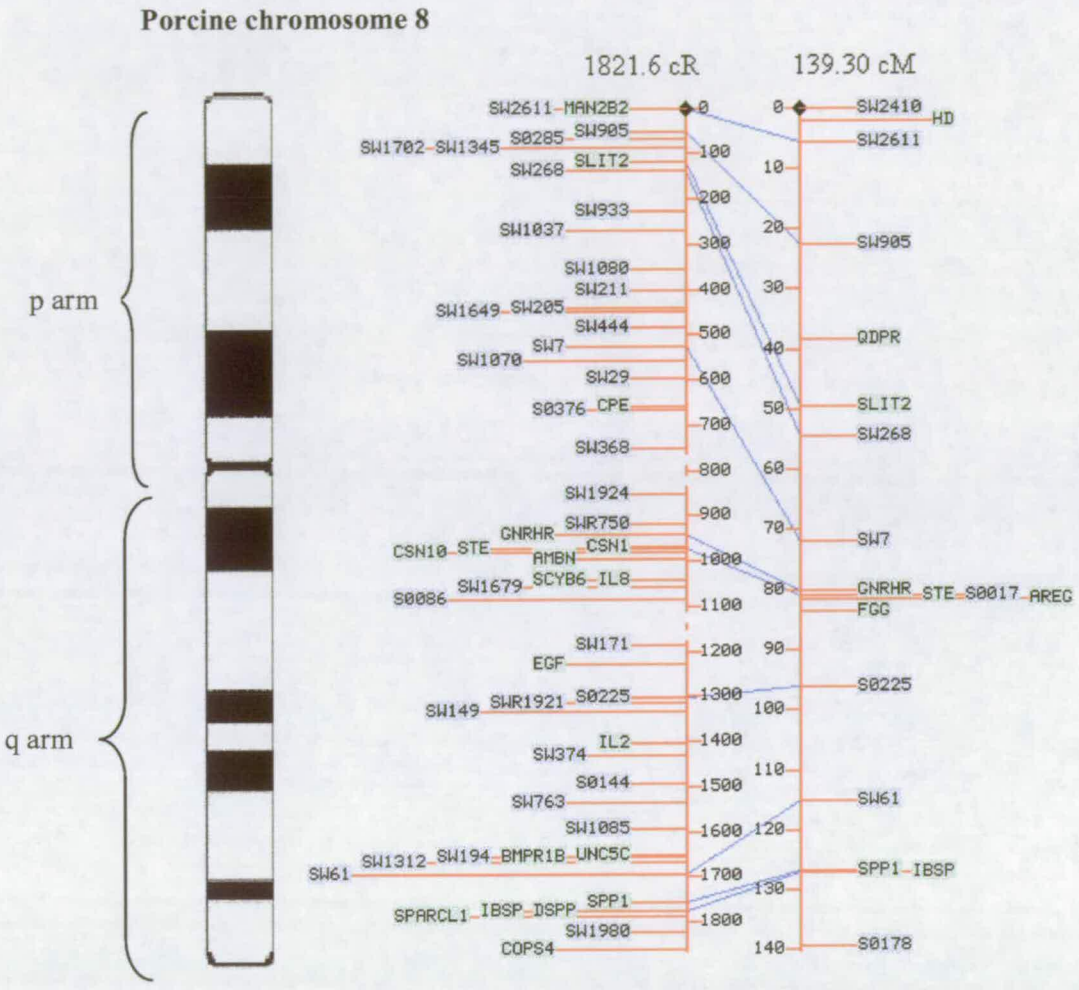


Figure 3-8 Radiation hybrid map (left) aligned with linkage map from reproduction QTL animals in chapter 2 (right), shown alongside the cytogenetic map of porcine chromosome 8. The maps of the three radiation hybrid linkage groups on SSC8 are separated by a dotted line, this distance between these groups is unknown. Markers highlighted in green are associated with known genes and those highlighted in blue are anonymous DNA markers.

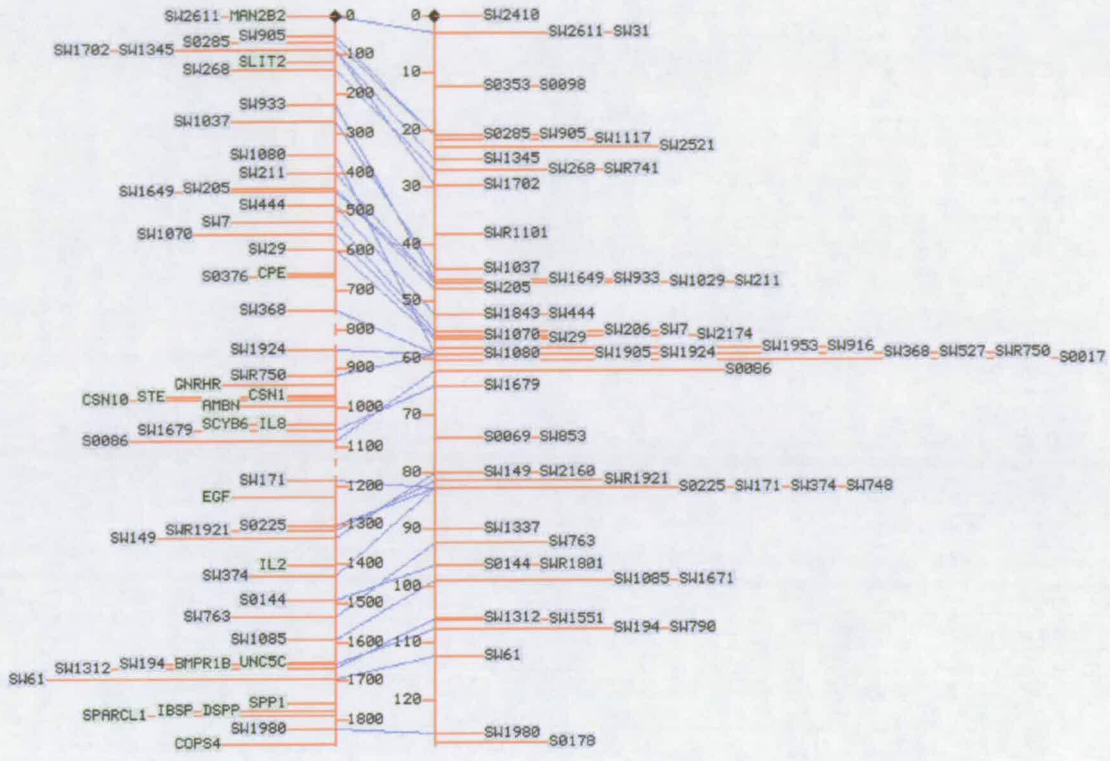


Figure 3-9 Radiation hybrid map (left) aligned with linkage map from USDA-MARC (right). The maps of the three linkage groups of the RH map of SSC8 are separated by a dotted line and this distance between groups is unknown. Markers highlighted in green are associated with known genes and those highlighted in blue are anonymous DNA markers.

3.3.5. Marker retention frequencies over radiation hybrid panel

The diploid marker retention frequency for each marker was then plotted against the position of that marker along the chromosome to determine whether the frequencies were indeed random along the length of the chromosome. Using Minitab™ (release 13.32) scatter graphs were drawn separately for the p and q arm of SSC8 and a smooth line drawn through the two sets of points (Figure 3-10). The function used was "LOWESS", or LOcally-Weighted scatterplot smoother, which shows the general relationship between the x and y variables. It must be noted that it is not a line of best fit though the data points. It can clearly be seen that the marker retention frequencies are not random along the chromosome.

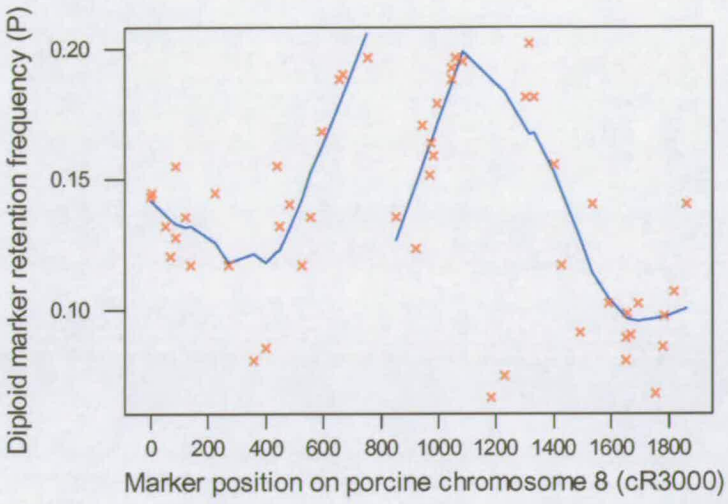


Figure 3-10 Diploid marker retention frequency along the p arm and q arm of porcine chromosome 8

3.4. Discussion

The mapping of polymorphic and non-polymorphic markers from various sources to porcine chromosome 8 was successful. Of those markers that were successfully pre-screened, 80 % were mapped to SSC8. In order to achieve this relatively high success rate, it was realised that the PCR conditions must be optimised to give a single bright band on the radiation hybrid pre-screen gel, to be most confident that the subsequent screen test will work.

It was also noted that the scoring of positive and negative clones for the screen test was dependant on a degree of subjectivity. It had been decided beforehand that all visible bands, whether faint or bright, would be scored and it was subsequently decided that it is important to have the same operator or pair of operators scoring the screen results throughout the whole experiment to ensure continuity for the subsequent analyses.

A problem of producing hybrids from a diploid cell line is that where a marker is seen to be present within a clone, it is not possible to know whether it is present as a single copy or as two copies within the same hybrid (Lunetta et al., 1995). The animal used to create the Cambridge-Roslin radiation hybrid panel used for this experiment was a male F1 Large White x Meishan. Sometimes the two alleles of different breed origin for each marker could be seen by the gel separation to be different sizes (e.g. Figure 3-5 shows a small difference in marker amplification size between the different positive clones). In addition, with a couple of the gels for the screen test (e.g. for marker *SW1085*) the positive clones amplified bands of several different sizes. This must demonstrate non-specific binding of the primers to additional sites other than the intended marker on SSC8. Therefore the screen test was only considered to have worked, where no more than two fragments of different sizes were amplified and that those different sizes were relatively close together.

CarthaGene, the program used to construct the radiation hybrid map, was developed by Thomas Schiex, Christine Gaspin *et al* from INRA (Institut National de la Recherche Agronomique) in Toulouse. Unlike other programs used to analyse RH

data, CarthaGene combines local two-point searches with multipoint maximum likelihood criteria, it can cope with errors in the data or missing values (i.e. up to three discrepancies in the typings between duplicate plates) and is computationally efficient (Schiex *et al.*, 2001 and (Schiex and Gaspin, 1997).

The approach used to interpret the RH data is a parametric method of maximum likelihood estimation (MLE), which assumes equal retention rates for all markers and utilises an Expectation Maximisation (EM) algorithm. This analysis model is a good compromise between computational efficiency and realism in the data output (Schiex and Gaspin, 1997). The same type of model is used as the basis for the programs RHMAP and RHMAPPER, however the improved version of the model used for CarthaGene is around twenty times faster to run than, for example, the program RHMAP 3.0 (Schiex et al., 2001).

Due to the number of markers that were mapped to SSC8 it was not possible to calculate the distance between the three separate linkage groups, in particular the centromeric region separating the p and q arms. In order to be certain of the distances between the groups, more markers would need to be mapped to these gap regions. It was not necessary to construct a complete map of SSC8 as the aim of this experiment was to map several genes in the identified QTL regions using a framework microsatellite map and not to build a high-density radiation hybrid map of this chromosome. By aligning the maps of the three linkage groups with the meiotic map built for the QTL analysis, it was possible to position the groups relative to one another. It can be seen from Figure 3-8 that the order of the eleven homologous markers on both maps is identical.

The maps built for the three linkage groups shown in order of their log likelihoods (Figure 3-7) can then be used to indicate how the marker order of the "best" map is supported by the data (Schiex and Gaspin, 1997). It can be seen from Figure 3-6 that for linkage group one, the LOD scores between pairs of markers for the first eight markers on the map (*MAN2B2* to *SW268*) are relatively high and Figure 3-7

shows that the order of these eight markers on the map is the same at all ten levels of log likelihood shown. It can also been seen that the distance between markers *SW933* and *SW268* is 0.9 Rays and the LOD score is lower than the threshold of 4.0, it is therefore not possible to be confident that these two markers map next to one another. Indeed the USDA linkage map in Figure 3-9 maps markers *SW1037* and *SW1702* between *SW268* and *SW933*.

The order determined by CarthaGene for the markers *SW211*, *SW1080*, *SW1037* and *SW933* in linkage group one is far from certain, in that the position of these markers varies at every level of log likelihood. In addition the LOD scores between each pair of markers is low especially between markers *SW1037* and *SW1080* where the LOD score is 2.2. The main difference in this region on the USDA map, is that *SW1080* is placed on the centromeric side of *SW29*. The order of the rest of the markers in linkage group one, from *SW205* to *SW368* is relatively certain at the four highest levels of log likelihood and diminishes below this. In addition the order of these markers is consistent with the USDA linkage map.

The likelihood of the order of the markers in linkage group two is relatively low. Only four of the markers also map to the USDA linkage map, two of which map in the same order. The location of the other two (*S0086* and *SW1679*) is reversed between maps.

In contrast, the likelihood of the order of the markers in linkage group three is high. The position of fourteen out of the twenty-one markers in this group is conserved for all ten maps at the varying levels of log likelihood. The order of markers *S0225*, *SWR1921* and *SW149* relative to one another is uncertain and indeed on the USDA linkage map the order of all three markers is inverted relative to the RH map (Figure 3-9). In the same way the order of markers *UNC5C*, *BMPR1B*, *SW194* and *SW1312* could not be resolved and the position of *SW1312* and *SW194* on the USDA map is also reversed. It appears that the more markers are mapped to a region (e.g. linkage group three), the more accurate the estimate of the marker order and marker distances.

The overall length of the RH map built of SSC8 was estimated to be at least 1900 cR_{3000} or 19 Rays. The distance between the three linkage groups is unknown and it is likely that a few additional so far unmapped markers will map to the far ends of the telomeres. McCarthy, L.C. *et al.* (unpublished) also built a RH map of SSC8 using a similar RH panel. They mapped 38 microsatellite markers and the total chromosome length was estimated to be around 2100 cR_{3000}. Yerle *et al.* (1998) built a RH map of SSC8 using a panel of 152 hybrids which had been irradiated at between 6000 and 7000 rads. They mapped 54 markers, the length of the map was around 2300 cR and the marker order was almost identical to that of the USDA linkage map.

The linkage map built by myself was around 140 cM and the equivalent map from the USDA was around 130 cM (see Figure 2-4). Schmitz *et al.* (1992) estimated the physical distance of SSC8 to be around 158 Mb. Therefore $1cR_{3000}$ on the 1900cR RH map roughly equates to 0.074 cM or 0.083 Mb (83 kb). Gyapay *et al.* (1996) demonstrated with a human RH panel created at 3000 rads, that 1cR was equivalent to around 300 kb. These two rough estimates of comparisons between physical and radiation hybrid distances for human and pig are similar. The estimate for the human is based on radiation hybrid maps of the whole genome, whereas the estimate for the pig is only based on a comparison between maps of chromosome 8 only. The explanation for the small difference in size is that recombination rates differ between chromosomes (Rohrer et al., 1996). The size of the human genome is believed to be similar to that of the pig (Archibald, 1994), the estimate of the human being 3200 Mb (Alberts et al., 2002).

By comparing the distance between markers on the genetic maps and on the radiation hybrid map, it is possible to compare recombination distances with actual physical distances on the chromosome. For example the distance between the adjacent markers *SW374* and *S0144* on my RH map is around 80 cR_{3000} (Figure 3-8) and on the RH map built by Yerle *et al.* (1998) the distance is around 270 $cR_{6000\text{-}7000}$. In comparison on the USDA genetic map the distance is relatively much larger at around 15 cM (Figure 3-9). If 1 cR_{3000} is equivalent to 0.074 cM, then the equivalent

genetic distance of 80 cR_{3000} would be predicted to be 5.9 cM; whereas it was in fact seen to be around 3 fold greater than this. Interestingly lower recombination rates have been observed between pericentric markers than telomeric markers (Rohrer et al., 1996). It appears from Figure 3-9 that the comparative distances between markers on the genetic and RH maps are indeed relatively similar around the centromere and the markers at the telomeres are relatively further apart on the genetic map than the RH map.

One of the assumptions of the CarthaGene analysis is that radiation induced breaks occur uniformly and at random across the chromosome (Slonim et al., 1997). Figure 3-10 clearly shows that the donor genome retention frequency varies across the chromosome, with the highest retention rates occurring in the pericentromeric regions and the telomeres. McCarthy (1996) showed that DNA around the centromere and to a lesser extent around the telomeres was retained at a higher rate than elsewhere on the chromosome. Chromosomes require a centromere and 2 telomeres to replicate and to be maintained within hybrid cells, these fragments are therefore retained at a higher rate (McCarthy, 1996). Cox *et al.* (1990), Kumlien *et al.* (1996), Raeymaekers *et al.* (1995) and Benham *et al.* (1989) also demonstrated non-random retention frequencies across the chromosome. They all describe higher retention rates around the centromeres and telomeres.

Many models of RH mapping assume equal marker retention frequency (e.g. WEBMAP (Newell et al., 1998), RHMAP (Walter et al., 1994) and CarthaGene (Schiex and Gaspin, 1997)), however Boehnke *et al.* (1991) and Lunetta and Boehnke (1994) propose a model that takes into consideration the centromeric or telomeric effect on fragment retention. Lunetta and Boehnke (1994) utilised a data set where the retention was greatest around the centromere, to compare a centromeric model to the equal retention model. They state that where the equal retention model was used, the resulting map length was underestimated. Therefore it would be ideal to run an analysis that included a model that allowed for the centromeric and telomeric effects of marker retention frequency, where variation has been shown to occur along the length of the chromosome.

In this experiment, the mean marker retention frequency for porcine chromosome 8 was calculated to be 13.5 % i.e. any one specific marker on the chromosome will be retained on average in around 13 out of the 94 hybrid clones. As a comparison, the average chromosome-specific retention rate for Research Genetics diploid hybrids is around 16 % per hybrid (Lunetta et al., 1995).

Although the radiation hybrid map built of porcine chromosome 8 can only be considered to be a draft map with respect to the accuracy of the approximation of distances between markers, the estimates of marker order is relatively certain. The map built can therefore be confidently used to identify additional genes, which have been predicted from the comparative mapping information of human chromosome 4, to map to the relevant QTL regions on porcine chromosome 8.

Chapter Four

4. USE OF COMPARATIVE MAPPING TO IDENTIFY POTENTIAL CANDIDATE GENES IN QTL REGIONS

4.1. Introduction

Comparative genome mapping not only gives us a much wider understanding of the history of chromosomal evolution. It is also a valuable tool for the transfer of genetic information from extensively studied "map-rich" species such as humans and mice to less studied "map-poor" species in particular livestock (O'Brien et al., 1993).

Traditionally comparative genomics was based on direct comparisons of gene sequences. However recently genome projects in several species have resulted in the development of detailed gene maps and genomic sequences which now allow whole genomes comparisons. The genome size, number of chromosomes and genes, the order of genes along the chromosome and the abundance and size of introns and repetitive DNA differs greatly among organisms (Alberts et al., 2002). However the conservation of gene order specifically in mammals of Eutherian order is high (Bishop, 2000) and where large chromosomal blocks are conserved across species it is possible to confidently predict where genes of interest will map from one species to another.

An exchange of genetic information between humans and other animals will bring major benefits to the study of the organisation and function of the human genome, as well as offer many practical advantages for animal breeding studies, in particular for agricultural traits of economic interest (Andersson et al., 1996). For example quantitative trait loci (QTL) for traits such as body weight, fat content, reproductive performance and disease susceptibility are likely to be conserved across mammals. Where these loci have been extensively studied and mapped in animal studies, it is likely that this knowledge can be used to help our understanding of abnormal growth and development in humans, where large well structured population studies are not possible (Bishop, 2000).

An example of this exchange of genetic information followed the identification of a mutation in the ryanodine receptor gene (*RYR1*) as the cause of malignant hyperthermia in pigs (Fujii et al., 1991). In pigs susceptibility to malignant hyperthermia was shown to be controlled by a recessive gene at a single autosomal locus mapped to SSC6 (Davies et al., 1988). The ryanodine receptor gene was mapped to human chromosome 19 and a single base transition discovered within the gene. This mutation was correlated with some inherited forms of malignant hyperthermia in humans (Gillard et al., 1991). The equivalent mutation in the porcine gene mapped to SSC6 was subsequently correlated to susceptibility to malignant hyperthermia in several breeds.

The number of genes in the genome is thought to be roughly correlated with the phenotypic complexity of that organism. For example the yeast *Saccharamyces cerevisae* has around 6,000 genes, *Drosophila melanogaster* around 13,000 genes and humans around 30,000 (Alberts et al., 2002). This predicted number of 30,000 protein-coding human genes was suggested after the draft of the genome sequence was released in 2001 (Lander et al., 2001) and many were surprised that the number was so low. However a few genes can have a large degree of complexity, due to alternative splicing and the many combinations available by switching genes on and off. It is likely that all genes found in complex organisms today have descended from a few ancestral genes from early life forms via gene duplication or mutation within a family of closely related genes, divergence and rearrangement of gene segments (Archibald, 1998).

The human and mouse genomes have been extensively studied. The draft sequence of the human genome was completed and made publicly available in 2001 by the international human genome sequencing consortium (Lander et al., 2001). Then in 1999 an international effort to sequence the genome of the inbred stain of mouse C57BL/6J was started. On May 6, 2002 a draft sequence was assembled by the Mouse Genome Sequencing Consortium, an international team of researchers from the Whitehead Institute in Cambridge, MA, Washington University in St. Louis, MO, and the Wellcome Trust Sanger Institute and the European Bioinformatics Institute,

in Hinxton, England (http://www-genome.wi.mit.edu/media/press/pr_mousegenome.html). This complete draft sequence was recently made publicly available (Waterston et al., 2002) and initial comparative analyses were carried out between human and mouse.

The size and the number of genes of the mouse and human genomes are similar. However it is around one hundred million years since the two species diverged. Due to the shorter generation interval, mice have had a relatively higher mutation rate in this time than humans; therefore there has been extensive sequence divergence in non-functional regions of the genome. Local gene order and the organisation of the chromosomes are very different, such that there have been around 180 breakage and rejoining events since the two lineages shared a common ancestor (Alberts et al., 2002). Interestingly Burt *et al.* (1999) predict that the chicken genome is more conserved with respect to the human genome than with the mouse genome. They predicted 96 conserved segments between the human and chicken genomes and 152 between mouse and chicken.

In contrast the human genome is surprisingly similar to that of the domestic farm animals, in particular the pig (*Sus scrofa*). Rettenberger *et al.* (1995) utilised the technique of heterologous chromosome painting or ZOO-FISH, where porcine chromosomes are "painted" with human probes from all chromosomes. This revealed that the 22 human autosomes were conserved in 45 porcine syntenic segments and that four chromosome pairs, including human chromosome 4 (HSA4) and porcine chromosome 8 (SSC8), appeared to be entirely conserved. Although large blocks of chromosomal regions do appear to be conserved across the two species, the gene order within these segments is unlikely to be preserved. However it is not until more detailed gene mapping information is available in livestock species that the precise level of conservation can be determined. Obviously the fewer the number of comparatively mapped genes between species, the larger the conserved blocks will appear.

Cargill *et al.* (1998), Ellegren *et al.* (1993a), Johansson *et al.* (1995), Larsen *et al.* (1999), Rohrer (1999) and Jiang *et al.* (2002a) all studied the level of conservation between HSA4 and SSC8. These studies revealed that although SSC8 does appear to be entirely conserved with HSA4, the gene order clearly differs and indeed a small region of HSA4 does not map to SSC8.

The aim of this chapter has been to utilise this high level of conservation between human chromosome 4 and pig chromosome 8 to attempt to comparatively map several genes from the homologous region on HSA4 to the regions of the two QTL, which were significant at the chromosome wide level for age group one sows. These were the teat number QTL mapped near the centromere and the prenatal survival QTL mapped to the telomere of the q arm of SSC8 (Table 2-4). This would give a better understanding of the level of conservation of synteny and of gene order in the particular regions of interest. It was therefore hoped that potential physiological candidate genes would be comparatively mapped to the region of the QTL. If this was not possible, then at least if the precise detail of how this region is conserved between the two species was known; then it would be possible to confidently predict which of the many mapped human genes would contain porcine homologues in the particular regions of interest on SSC8.

The potential role of the pig homologues of these human genes in controlling embryo survival rates, from a maternal point of view could then be investigated. However the physiology of pregnancy, in particular the type of placenta formed and the number of offspring, is very different between humans and pigs and this would have to be taken into consideration. Another important outcome of this study is that it will give us a better understanding about the general level of gene order conservation and the patterns of re-arrangement between human chromosome 4 and pig chromosome 8.

4.2. Materials and Methods

4.2.1. Identification of homologous porcine genomic or EST sequences to genes on human chromosome 4

A detailed gene map of human chromosome 4 was obtained from the Online Mendelian Inheritance in Man (OMIM) database from the NCBI website (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM). As previously mentioned human chromosome 4 is extensively conserved with porcine chromosome 8 and the 4q11-q28 region is believed to be homologous to the q arm of porcine chromosome 8. The telomeric section of the q arm is where the prenatal survival and litter size QTL have been mapped (Figure 2-6C).

To achieve a rough estimate of the location of the homologous region on human chromosome 4 (HSA4) to the q arm of SSC8, the position of the genes previously mapped to SSC8q (from Figure 2-4) was aligned with the physical map of HSA4. Figure 4-1 indicates and Ellegren *et al.* (1993a) suggested that there was a breakpoint and inversion event around the site of the QTL, such that the equivalent area on HSA4 to the telomeric region of SSC8 q arm can be found near the centromere. The area of selection was extended away from the predicted QTL region to incorporate from the centromere on HSA4 to near the *FGG* gene in region 4q28 (where the inversion appears to have occurred). This was done to be certain that the whole region of interest around the QTL would be selected. It was also hoped that this would allow the exact site of the inversion to be defined with more precision and to determine whether the homologous region to the QTL was indeed conserved in the human. Using the GeneCards™ program found at http://bioinfo.weizmann.ac.il/cards/index.html the GenBank accession number was identified for every known gene in the region 4q11-q28.

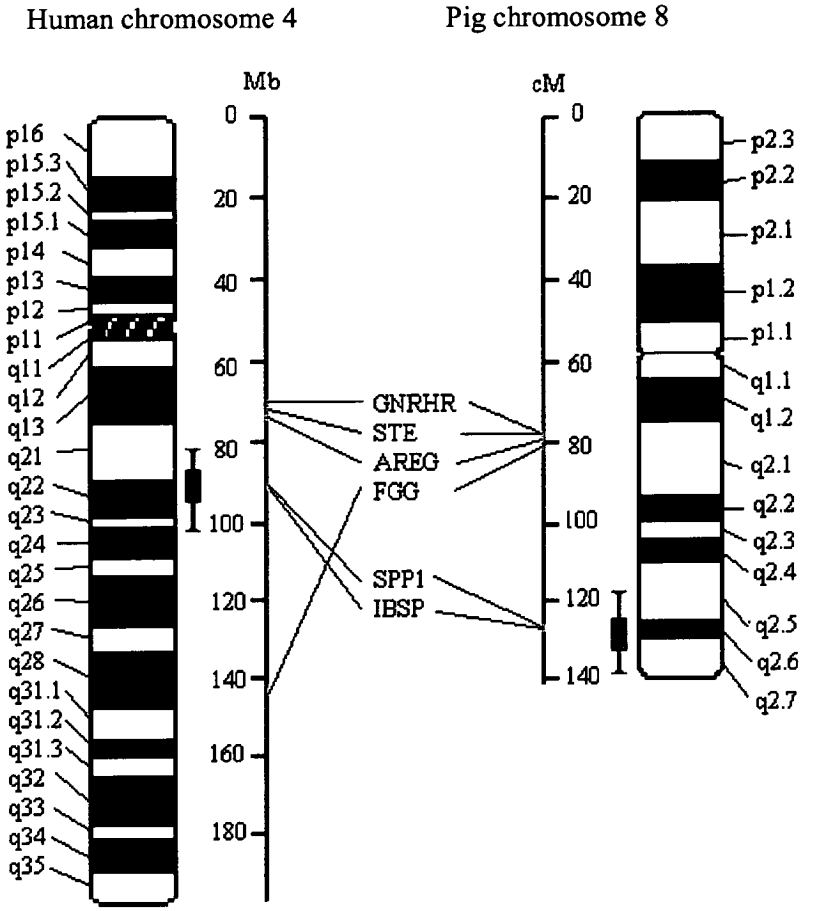


Figure 4-1 Diagrammatic representation of the relative position of the genes mapped to the q arm of SSC8 (from linkage map, Figure 2-4), on human chromosome 4. The approximate region of the prenatal survival and litter size QTL identified in chapter 2 is shown as a blue line, with the most significant region of the QTL marked as a solid box.

Using the standard nucleotide-nucleotide BLAST search for sequence homology (http://www.ncbi.nlm.nih.gov/BLAST/), the human gene sequences were searched for homology with genomic pig sequences in the "non-redundant" GenBank, EMBL, DDBJ and PDB databases, by selecting "*Sus scrofa*" as the selected organism. The sequence matches found are displayed in order of the level of conservation.

The BLAST search program was developed by Altschul *et al.* (1997) and involves the use of local pair wise alignments. An alignment score is produced by searching a "word size" of every 11 bases along the sequence of interest against all sequences present in the various databases mentioned. This therefore allows gaps to be introduced when matching the two sequences and an overall level of alignment ascertained. The output also reports a percentage identity between the porcine and human sequence. This can however be misleading because if an exact match to a very short porcine sequence is found, then the identity would be reported as 100 %, whereas in fact a large part of the human sequence would not have been matched to porcine sequence. It is therefore best to consider the alignment score, which is measured in "BITS", as the most reliable prediction of how accurate the match is. This BIT score is assigned an E value depending on how significant the match is. In order to avoid finding chance matches, a threshold of significance is produced from a model of random sequences. E values which are smaller than $1e^{-20}$ can be considered to be significant. As a rough rule of thumb this represents an alignment with a bit score greater than 80. Any alignment score less than this indicates that the homology with the pig sequence found is uncertain and it is likely that the match will be to a different region of genomic sequence than expected.

A draft RH map of these homologous porcine genes, genes known to map to SSC8 and microsatellites was built. Following this, additional genes were identified specifically in the homologous region of HSA4 to the prenatal survival QTL on SSC8 (between genes *BMP3* and *RAP1GDS1*). The human gene information was obtained from the "Markview" output of a physical map of HSA4 within the Ensembl website (http://www.ensembl.org/Homo_sapiens). At the present time there is a rapid input of human gene sequences into the reference databases and it

was believed that the Ensembl website would be the most up to date and accurate source of gene sequences mapped to HSA4. Several more genes in the region of the QTL were indeed identified.

The GenBank accession number was obtained for each gene and the sequence BLAST searched against EST as well as the "non redundant" porcine sequence databases. Any porcine EST sequences obtained were then checked to determine whether there were a cluster of EST sequences representing the gene or whether the EST was a singleton, for this the website for the TIGR institute of Genome Research was used (http://www.TIGR.org). Where clusters of ESTs are found, this increases the chance of successfully designing a marker within the gene of interest and not amplifying the wrong gene. TIGR have developed gene indices for species for which there are several thousand ESTs in the public domain.

4.2.2. Designing porcine gene markers to type over the radiation hybrid panel

Although heterologous probes and primers can be useful for gene mapping across species, pig specific primers were required for use with the hamster-porcine radiation hybrid panel. The RH panel consists of hybrid pig and hamster DNA and therefore by designing primers in less conserved non-coding regions of the gene, the chance of hamster DNA amplification is much reduced. Therefore the primers were designed in either the intronic genomic sequence, 5' untranslated region (UTR) or the 3' UTR of the porcine gene sequences. Where primers were designed within EST sequences, those that were a 3' read tended to be more successful than 5' reads; however unfortunately the majority of pig ESTs recorded are 5' reads.

The primer pair for each marker was designed to amplify product sizes of between 100 and 400 base pairs, an optimal size range for the PCR conditions to be used. Where a poly A tail was present at the 3' end of the sequence, this was removed prior to primer design. Primer design was carried out using the web based program Primer3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) (Rozen and Skaletsky, 2000). Where sufficient non-coding sequence was available,

two sets of independent primer pairs were designed from different regions of the porcine sequence that showed good homology to a human chromosome 4 gene. This was done to increase the chance that a marker in the gene of interest would be successfully amplified across the RH panel.

4.2.3. Constructing comparative maps across mammalian species

Map alignment of human chromosome 4 with pig chromosome 8 was done using the Anubis map viewer available at http://www.thearkdb.org/ (Hu et al., 2001). The physical map of human chromosome 4 from Ensembl was aligned with the radiation hybrid map of porcine chromosome 8 and also with the linkage map of SSC8.

Also the level of gene order conservation between human chromosome 4 and other species of interest, in particular livestock, was investigated. The chromosomal locations and gene orders for sheep and cattle were obtained from linkage maps of these two species found at http://www.thearkdb.org. The gene information for the mouse was obtained from a comparative map of human chromosome 4 with the mouse and the rat (http://www.well.ox.ac.uk/rat_mapping_resources). The gene information of human chromosome 4 and pig chromosome 8 is the same as that used to carry out the map alignment of these two species as outlined above.

4.3. Results

4.3.1. Markers used to build the initial draft RH map of SSC8

The genes in the human chromosome 4 region q11-q28 that showed homology to porcine genomic sequence are shown in Table 4-1. The sequence similarity score and the region of primer design within the porcine sequence and subsequent marker sizes are also shown. It was only possible to design primers to amplify markers across porcine sequence where sufficient non-coding sequence was available. In order to achieve successful mapping of the gene of interest, two pairs of independent primer sequences were designed, however with some shorter gene sequences it was only possible to design one pair. The primer sequences of those genes that were successfully screened over the radiation hybrid panel are shown in Table 3-2.

4.3.2. Markers used to increase the density of genes in the prenatal survival QTL region mapped to SSC8 qter

A list of around 300 genes and their physical map position in base pairs was obtained from the Ensembl website in April 2002. The particular region of interest for gene identification was the homologous area to the prenatal survival QTL region identified around the telomere of the q arm of SSC8. On human chromosome 4 this region was predicted to be between around 80 Mb and 100 Mb (see Figure 4-1).

All the genes that were known to map to that region at the time of the search of Ensembl are listed in Table 4-2. For the general search for homologous porcine sequences to human chromosome 4 genes carried out in section 4.3.1, only matches to porcine genomic DNA were considered in order to ensure that primers would be designed in non-coding gene regions. For this small chromosomal region of particular interest, it was considered worthwhile to also investigate the database of porcine EST sequences.

Table 4-1 Identity and BLAST score of human genes in the region 4q11-28 that were identified to display a significant match to porcine genomic DNA. Where sufficient non-coding sequence was available, the region of primer design is stated (CDS = coding sequence, UTR = untranslated region). Genes for which primer pairs were successfully designed are emphasised in red.

Human gene (GenBank accession no.)	Location	BLAST score (BITS)	E value	Porcine gene (GenBank accession no.)	Region of primer design (marker sizes)
ALB (M12523)	4q11-q13	155	1e-34	Liver albumin (M36787)	3'UTR (100bp)
KDR (AF035121)	4q12	438	1e-121	Flk-1 type VEGF receptor (AJ245446)	Only CDS
KIT (X06182)	4q12	1695	0	Mast/stem cell growth factor receptor (AJ223228)	Only CDS
IL8 (M17017)	4q12-13	182	1e-44	Interleukin 8 (M86923)	3' UTR (101bp and 208bp)
PPBP (M54995)	4q12-13	105	1e-21	Platelet basic protein (X77935)	Too little non-CDS available
SCYB6 (U81234)	4q12-21	133	1e-29	Alveolar macrophage-derived chemotactic factor II (M99368)	3'UTR (157bp and 198bp)
CSN10 (M73628)	4q13-21	123	6e-27	Kappa casein (X51977)	3'UTR (110bp and 151bp)
AMBN (AF219994)	4q21	601	1e-170	Ameloblastin (U43404)	3' UTR (198bp and 192bp)
IBSP (J05213)	4q21	190	4e-47	Bone sialoprotein (L10363)	3' UTR (197bp and 202bp)
SPP1 (AF052124)	4q21	202	1e-50	Secreted phosphoprotein 1 (X16575)	See *SPP1-1* Table 2-2
CSN1 (X78416)	4q21	80	9e-14	Alpha casein (X54973)	3'UTR (230bp and 185bp)
GNRHR (NM_000406)	4q21	559	1e-157	Gonadotrophin releasing hormone receptor (AF227686)	Intron (180bp and 189bp)
NKX6A (U66799)	4q21-22	90	3e-17	Homeobox protein NKX6.1 (AF236157)	Only CDS
PDHA2 (M86808)	4q22-23	888	0	Pyruvate dehydrogenase (X52990)	Only CDS
IL2 (J00264)	4q26-27	541	1e-152	Interleukin 2 (AB041341)	5'UTR (207bp and 101bp)
PMBP (XM_003419)	4q26	64	1e-8	Steroid membrane binding protein (X99714)	3' UTR (203bp and199bp)

(Note that the sequence for human *PMBP* (XM_003419) has since been withdrawn from the database – see section 3.3.4.1).

Table 4-2 Identity and BLAST score of human chromosome 4 genes in the region 80-100Mb aligned with porcine genomic DNA and ESTs. Genes for which primer pairs were successfully designed are emphasised in red.

Human gene (GenBank accession no.)	Position on HSA4 in bp (band)	BLAST score (BITS)	E value	Porcine gene (GenBank accession no.)	Region of primer design (marker sizes)
BMP3 (NM_001201)	81.8 (q21)			No match	
PRKG2 (NM_006259)	81.9 (q21)	519	1e-146	EST: 376238 (BI346879)	5' end of EST (101bp)
HNRPD (NM_031370)	83.0 (q21)	656	0	EST: IKU602768943.R1 (BM659693)	Within EST (180bp)
HNRPDL (NM_005463)	83.2 (q21)	502	1e-140	EST: MI-P-A2-acs-h-03-1-UM.s1 (BF711801)	3' end of EST (201bp)
COPS4 (NM_016129)	84.0 (q21)	793	0	EST: MI-P-A1-aao-d-06-1-UM.s1 (BF702060)	3' end of EST (206bp)
MRPS18C (AK002008)	84.3 (q21)	97.6	4e-19	EST: 300783 (BG382940)	5' end of EST (170bp)
NKX6A (NM_006168)	85.4 (q21)	422	1e-117	Homeobox protein NKX6.1 (AF236157)	CDS (198bp)
CDS1 (NM_001263)	85.5 (q21)			No match	
MAPK10 (NM_002753)	87.0 (q21)	795	0	EST: 389634 (BI467831)	5' end of EST (114bp)
MLLT2 (NM_005935)	88.0 (q22)	268	8e-70	EST: BJP602768611.R1 (BM659873)	Within EST (197bp)
SPARCL1	88.3 (q22)			See Table 3-2	
DSPP (NM_014208)	88.4 (q22)	184	1e-44	Dentin sialoprotein (AF332578)	3' UTR (112bp)
DMP1 (NM_004407)	88.4 (q22)			No match	
IBSP	88.6 (q22)			See Table 4-1	
MEPE (NM_020203)	88.6 (q22)			No match	
PKD2 (AF309082)	88.8 (q22)	708	0	EST: 223850 (BF075079)	5' end of EST (171bp)
SPP1	89.0 (q22)			*SPP1-1* see Table 2-2	
NEK1 (AL050385)	89.1 (q22)	149	3e-34	EST: SSK27 (X91312)	3' end of EST (121bp)

Table 4-2 ctd

Human gene (GenBank accession no.)	Position on HSA4 in bp (band)	Homology score (BITS)	E value	Porcine gene (GenBank accession no.)	Region of primer design (marker sizes)
HERC3 (NM_014606)	89.6 (q22)	811	0	EST: 362476 (BI338632)	5' end of EST (182bp)
SNCA (NM_000345)	90.9 (q22)	129	8e-29	EST: 76891 (AW436578)	Within EST (165bp)
MMRN (NM_007351)	91.0 (q22)			No match	
GRID2 (NM_001510)	94.5 (q22)			No match	
PGDS (NM_014485)	95.5 (q22)			No match	
BMPR1B (NM_001203)	96.3 (q22)	2129	0	Bone morphogenetic protein type 1 receptor (AY065994)	3'UTR (207bp) and Intron 6* (186bp)
UNC5C (NM_003728)	96.4 (q22)	505	1e-142	EST: UNL-P-FN-bh-a-07-o-UNL (BI182276)	3' end of EST (208bp)
PDHA2 (NM_005390)	97.2 (q22)	888	0	Pyruvate dehydrogenase (X52990)	CDS (111bp)
RAP1GDS1 (NM_021159)	99.1 (q22)			No match	

* Sequences for a primer pair in intron 6 of porcine *BMPR1B* obtained from Gary Rohrer (USDA-MARC).

An EST (Expressed Sequence Tag) is a partial, single pass sequence from either end of a cDNA clone and therefore represents a small part of the coding region of a gene. They can be used as tags to identify specific genes throughout the genome mainly in functional genomics studies involving gene expression studies in particular tissues (Adams et al., 1991). As mentioned previously, it is highly likely that markers designed from conserved gene regions (i.e. coding sequences) may prove difficult to map on the RH panel as the hamster homologue may well also be amplified. However in order to identify potential candidate genes for the control of prenatal survival it was necessary to map as many genes as possible in the QTL region. I needed to be certain that this region was completely conserved in the human to be able to confidently predict additional genes from the human data that would map to the region on porcine chromosome 8. Therefore it was worthwhile attempting to amplify markers within the RH panel that were designed from EST sequences.

Table 4-2 contains the identities and GenBank accession numbers of all the genes that map in the region between 80Mb and 100Mb on human chromosome 4. The porcine EST or genomic sequences, which were shown to match to these human genes, are also shown, along with the region within the porcine DNA sequence that the primer pairs were designed and the size of the markers.

4.3.3. Comparative maps of HSA4 and SSC8

Porcine markers were successfully designed for the homologues of 24 human chromosome 4 genes that were predicted to and had previously not been mapped to pig chromosome 8 (*ALB, IL8, SCYB6, CSN10, AMBN, CSN1, IL2, PMBP, PRKG2, HNRPD, HNRPDL, COPS4, MRPS18C, NKX6A, MAPK10, MLLT2, DSPP, PKD2, NEK1, HERC3, SNCA, BMPR1B, UNC5C* and *PDHA2*).

It can be seen from Table 3-2 that twelve of these markers were successfully screened over the radiation hybrid panel i.e. they did not amplify hamster DNA. Ten out of these twelve markers were mapped on the RH map built of porcine chromosome 8 (see Figure 4-2). The RH map built for SSC8 was aligned with the

physical map of HSA4 obtained from Ensembl (Figure 4-2, the ten genes mapped comparatively are outlined in pink). All the genes present between 80 Mb and 100 Mb on HSA4, the predicted homologous area to the prenatal survival QTL region, are shown, whereas only a few genes on the rest of the chromosome are shown, in order to prevent the map being too cluttered. Also a few genes had been placed on the linkage map (from Figure 2-4) and not on the radiation hybrid map, therefore in order to define the region of homology with more clarity the linkage map produced in chapter 2 was also aligned with the gene map of human chromosome 4 (Figure 4-3).

Figure 4-4 shows a diagrammatic representation of the level of gene order conservation between SSC8 and HSA4. It can clearly be seen that the region from *COPS4* to *BMPR1B* (i.e. the area surrounding and the prenatal survival QTL region itself) is a conserved synteny group. The order of the genes within the group is conserved relative to one another, however the whole group has been inverted. Initially it would appear as if the whole region from *FGG* to *COPS4* has been inverted however the genes *EGF* and *IL2* were seen to be in the same order relative to each other. For the exact level of homology in the region to be resolved, more genes need to be mapped between *FGG* and *BMPR1B*. The gene order appears to be entirely conserved from the telomere of the p arm on both chromosomes to around the gene *AREG* (around the centromere). The only exception is the translocation of a small segment around *CPE* from the q arm of HSA4 to the p arm of SSC8. The end of the q arm of HSA4 does not map to SSC8.

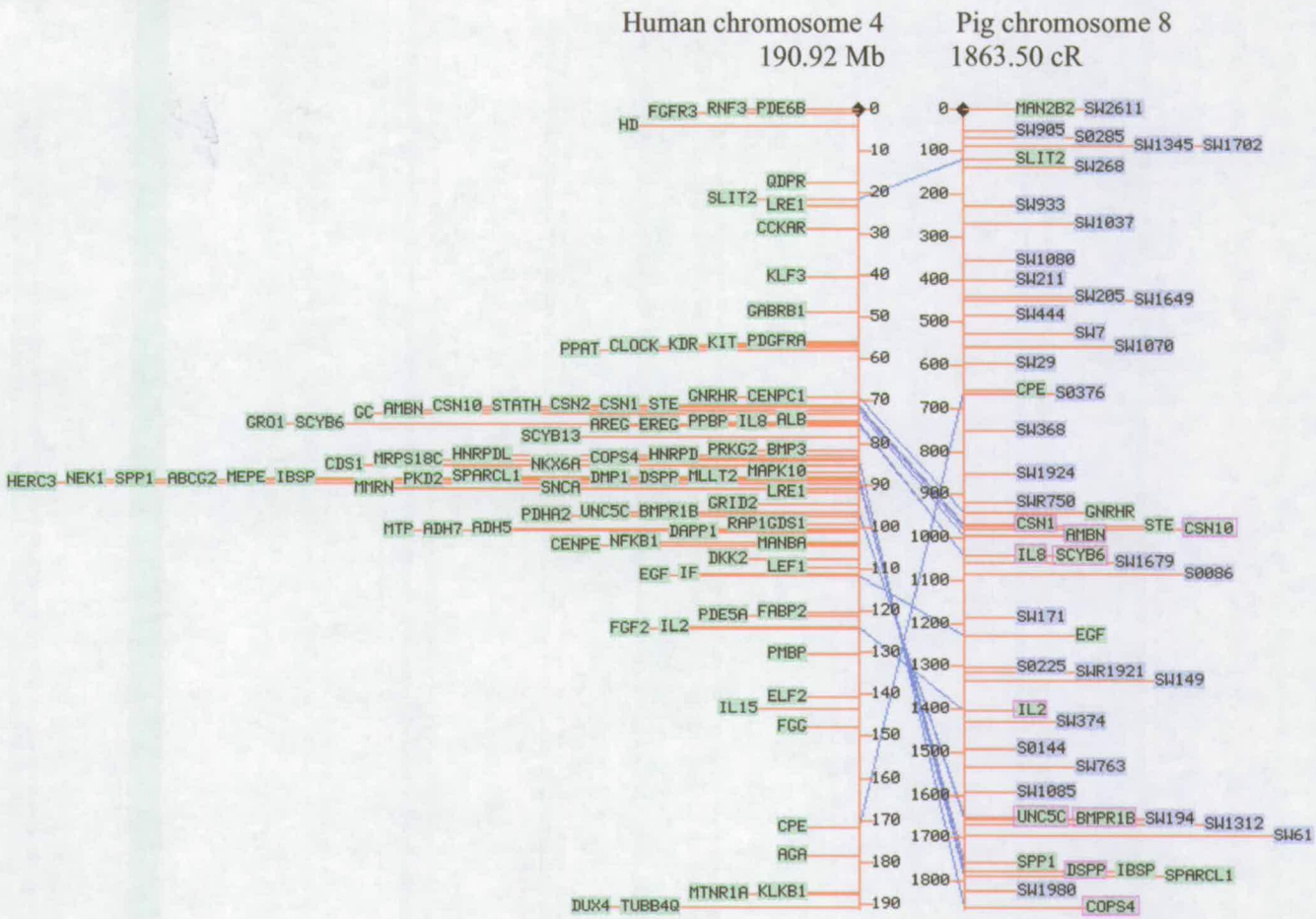


Figure 4-2 The radiation hybrid map built of porcine chromosome 8 (from Figure 3-8) aligned with the physical map of human chromosome 4 obtained from Ensembl (10 comparatively mapped genes are outlined in pink).

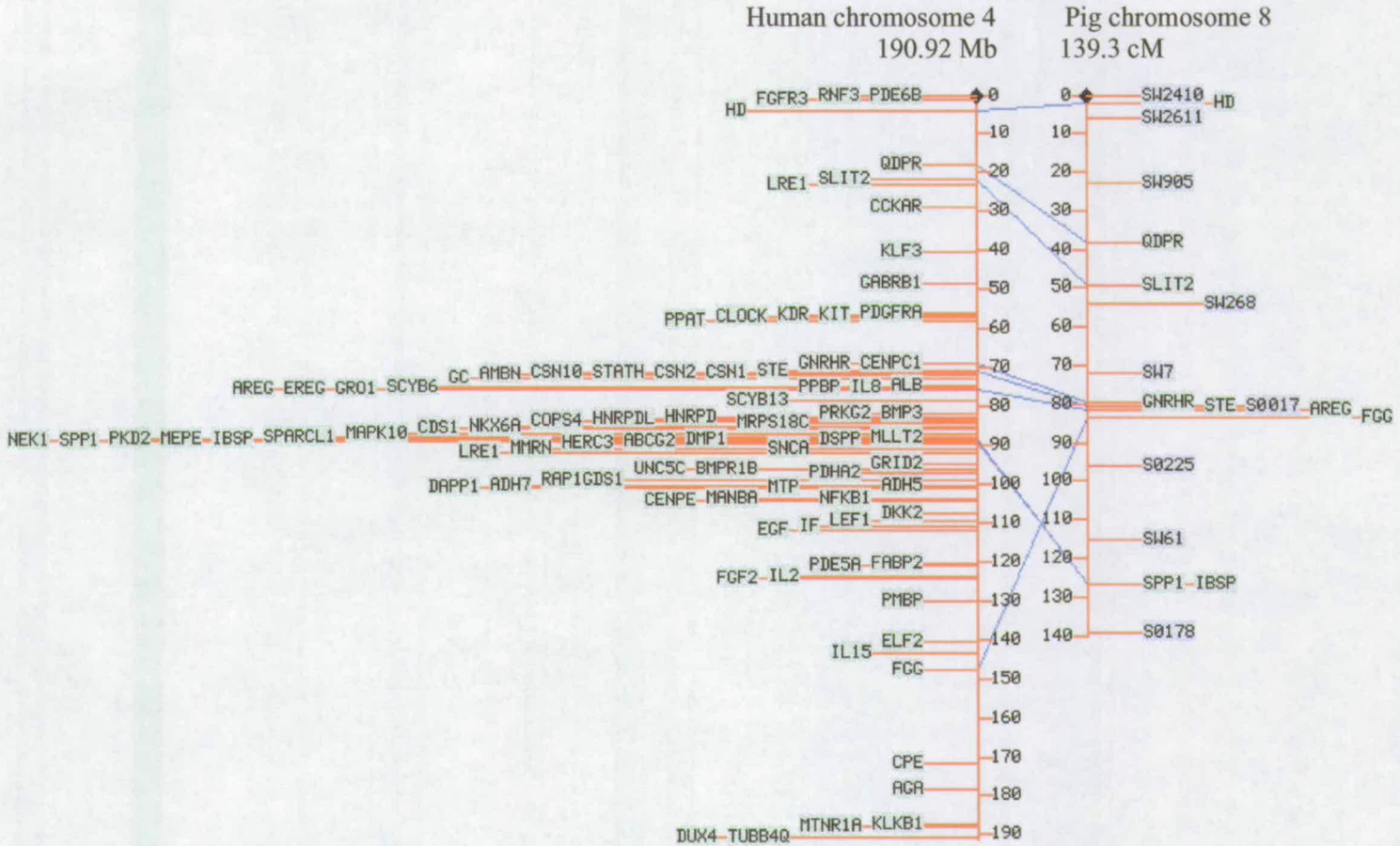


Figure 4-3 The linkage map built of porcine chromosome 8 (from Figure 2-4) aligned with the physical map of human chromosome 4 obtained from Ensembl.

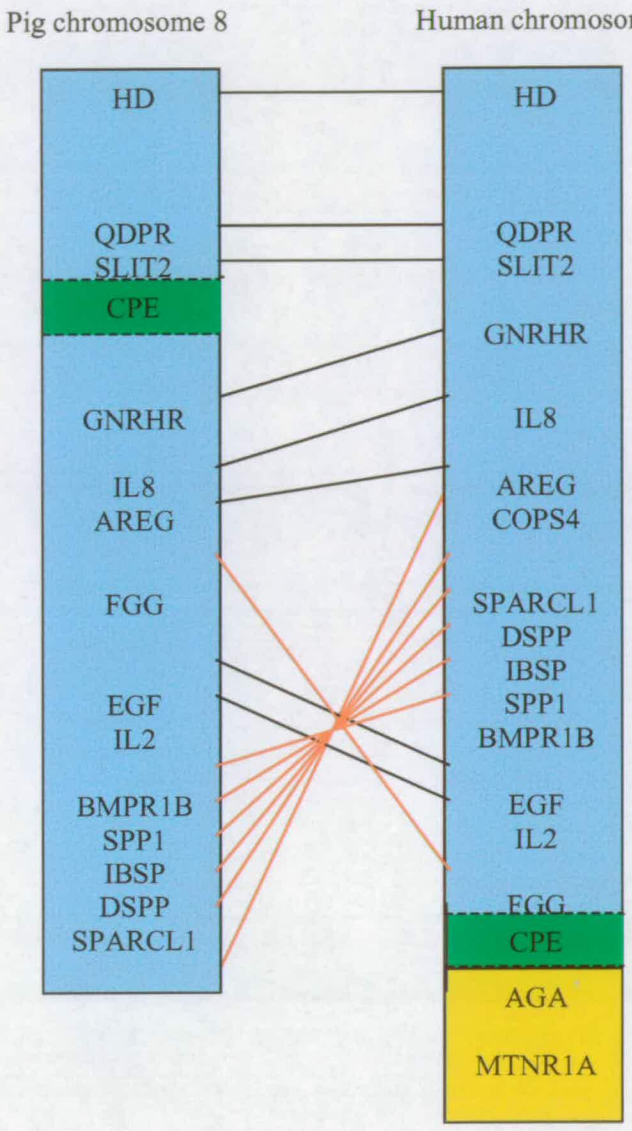


Figure 4-4 Representation of the conservation between HSA4 and SSC8. Major inversion event (red lines) and relocation of small segment (green) highlighted. Orange section of HSA4 is not homologous to SSC8. Key gene markers are labelled.

4.3.4. Comparative maps of HSA4 with other mammalian species

Figure 4-5 is a simplified representation of the level of gene order conservation between human chromosome 4 and the pig, mouse, sheep and cattle genomes. Only those genes whose positions are known in two or more species are shown. As the position of many genes is unknown the level of conservation across species, especially for sheep and cattle, will be more complex than is shown here. As mentioned previously porcine chromosome 8 is homologous to almost the whole of human chromosome 4. There is evidence of a region of inversion from *BMPR1B* to *COPS4* and a region that has been relocated around *CPE*. The telomere of the q arm of HSA4 appears to be least conserved. *AGA* maps to pig chromosome 15 (Jiang *et al.*, 2002a and (Larsen et al., 1999) and *MTNR1A* to pig chromosome 17 (Jiang *et al.*, 2002a; Messer *et al.*, 1997 and (Larsen et al., 1999). The order of the homologous genes on mouse chromosome 5 appears to be identical to that of the equivalent region on human chromosome 4, except for the very telomeric region of the p arm of HSA4 (around *PDE6B*), which appears to have broken off and rejoined to the far end of mouse chromosome 5. The rest of HSA4 shows a very low level of conservation in the mouse.

Gene order conservation appears to be very similar between cattle and sheep and as with the comparison of mouse and human, the lower half of HSA4 q arm is less well conserved with these two species than the rest of this chromosome. Interestingly the order of sheep and cattle genes in the block highlighted in blue on Figure 4-5 is very different to that of the pig, human or mouse.

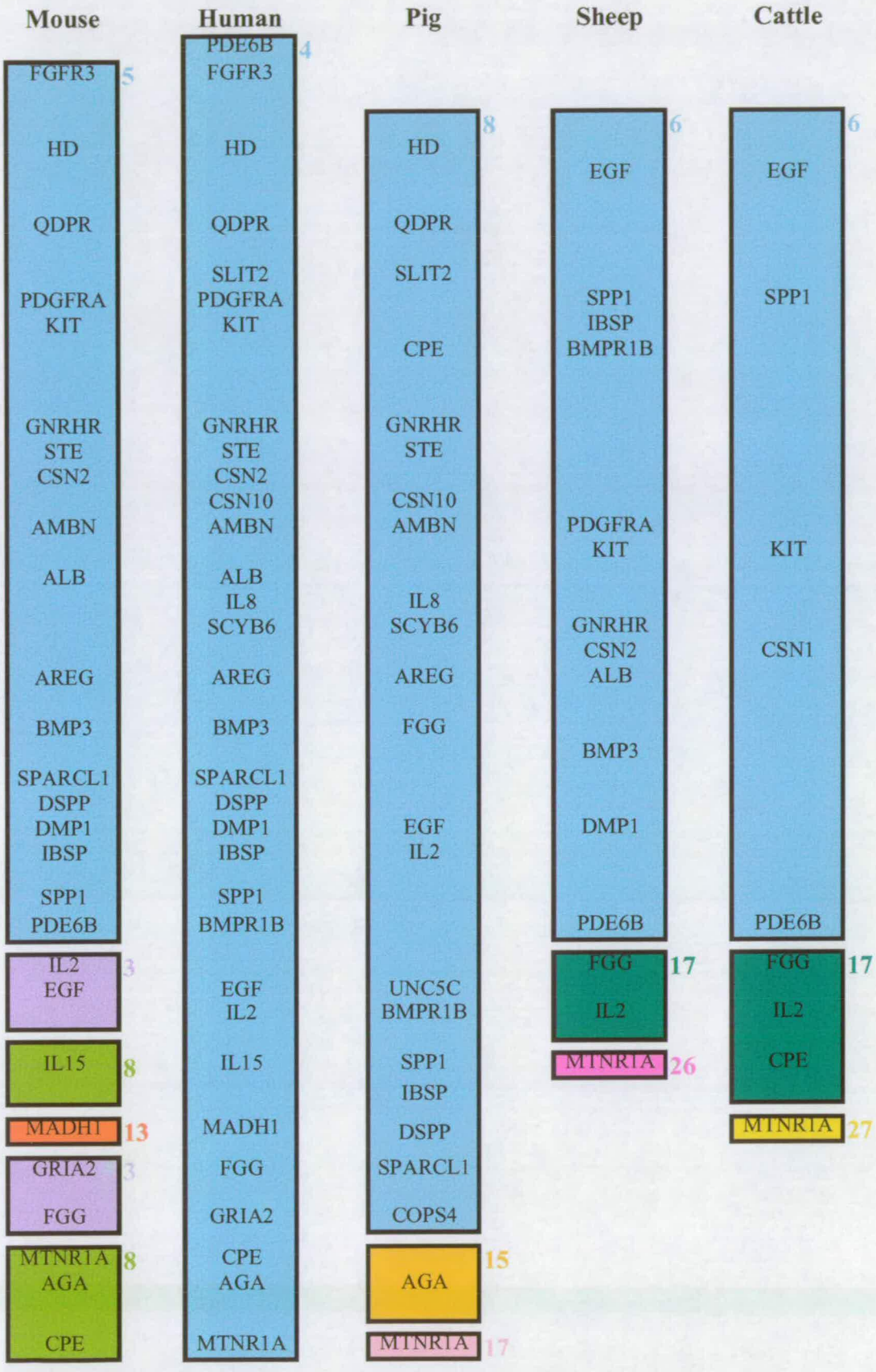


Figure 4-5 Comparative mapping of human chromosome 4 with pig, mouse, sheep and cattle. Only key genes that map across more than one of the species are shown (The position of several genes is unknown). The chromosome numbers are marked next to each coloured block.

4.4. Discussion

The alignment of the detailed gene map of human chromosome 4 with the radiation hybrid map built of pig chromosome 8 revealed that although the level of conservation of synteny between these two chromosomes is extensive, the conservation of gene order is more complex than previously thought. Ellegren *et al.* (1993a) carried out one of the early porcine gene mapping studies and they mapped the genes *IL2, ADH3* and *SPP1* to porcine chromosome 8 relative to the gene markers *PDGFRA* and *ALB*. They resolved the gene order as *PDGFRA, ALB, IL2, ADH3* and *SPP1*. They state that this group of genes, as well as *FGA, FGB* and *FGG* linked to *ALB* by Archibald *et al.* (1992), cover around three quarters of the q arm of HSA4, one of the longest chromosome arms in the human (Morton, 1991). Ellegren *et al.*, (1993a) believe that the fact that these loci are syntenic in many primates and other mammals such as the pig, suggests that the group was present together on a single ancestral mammalian chromosome.

Rettenberger *et al.* (1995) hybridised chromosome specific probes, representing each of the human autosomes, to pig chromosomes in metaphase state in order to determine the level of synteny between the human and pig genome. They describe that human chromosome 4 is conserved as the whole of porcine chromosome 8. Johansson *et al.* (1995) built a porcine comparative map of 38 genes whose presumed homologues had been mapped in humans and/or mice. They showed that, as would be expected, there are significantly more interchromosomal rearrangements between the pig and mouse genomes (around 77) than between pig and human genomes (around 35).

Johansson *et al.* (1995) describe how the linkage group *PDGFRA, ALB, SPP1, ADH3, IL2, FGG* on human chromosome 4 has been disrupted by two independent translocations to mouse chromosomes 5 and 3 and to cattle chromosomes 6 and 17. Whereas the entire group is found on SSC8, but with at least one intrachromosomal rearrangement of gene order between *FGG* and *SPP1*, which appears to be unique to the pig lineage. As a result of the comparative mapping study described in this

chapter more genes have been mapped to SSC8 and the gene order resolved in more detail.

The homologous region to the telomere of the q arm of HSA4 does not map to SSC8. *AGA* and *KLK3* map to HSA4q35 and to pig chromosome 15 (Jiang *et al.*, 2002a and (Larsen et al., 1999) and *MTNR1A* maps to the distal region of HSA4q35 and to pig chromosome 17 (Jiang *et al.*, 2002a; Messer *et al.*, 1997 and Larsen *et al.*, 1999). Larsen *et al.* (1999) suggest that there appears to be an increased level of rearrangements around the telomeric regions of chromosomes across mammalian species. This was certainly seen in Figure 4-5, when comparing the telomere of the q arm of HSA4 with mouse, pig, sheep and cattle genomes. One explanation for this phenomenon could be that there is an increase in meiotic recombination rate around the telomeres (Larsen et al., 1999).

The detailed gene map of HSA4 from Ensembl, showed that the most telomeric gene on the p arm appears to be *PDE6B*. This gene was not mapped in this study and it was therefore not possible to be certain whether the whole of HSA4p was conserved on SCC8. However Rohrer (1999) did map *PDE6B* and not unsurprisingly it was located at the telomere of the p arm of SSC8. Therefore it appears that the whole section of HSA4 from *PDE6B* to *CPE* is conserved in SSC8, but with some rearrangement of gene order (see Figure 4-5).

Out of the twenty-four genes on HSA4 that revealed homology to published porcine genomic or EST sequences in this study, only ten were successfully mapped to SSC8. This clearly demonstrates the limitation of using radiation hybrid mapping to map gene sequences, especially where only coding sequence is available for primer design. Only two of the porcine EST sequences were successfully used to map these genes to SSC8. ESTs are the most readily available type of gene sequence information stored in the databases; there are currently 110,437 porcine EST sequences in the GenBank database (November, 2002). However around 70 % of these ESTs are 5' reads and as mentioned previously 5' reads are much more likely to contain coding sequence than 3' reads and the primers used will most likely

amplify the homologous gene from the hamster DNA in the hybrid RH panel. This lack of success mapping EST sequences is unfortunate because radiation hybrid mapping is a relatively straightforward technique and could allow many markers to be mapped much more rapidly than would be possible with linkage map construction.

The radiation hybrid panels are constructed from hybrid mammal and rodent DNA, two groups of animals that are relatively distant in evolutionary terms. Although most of the genome structure and sequence will be very different, the coding sequences of homologous genes are still highly conserved. This same problem was encountered following the development of comparative anchor tagged sequences (CATS) by Lyons *et al.* (1994), and Traced Orthologous Amplified Sequence Tags (TOASTs) by Jiang *et al.* (1998). The idea of both these studies was to aid the establishment of a common framework map for locating homologous genes across several mammalian species, by the design of "conserved" primer pairs within these gene sequences. However for the same reason that the gene marker sequences are conserved across species, these tags were not successfully mapped by radiation hybrid mapping. Jiang *et al.* (1998) suggest either refining the PCR conditions to advantage the donor species or to create a hybrid panel using more evolutionary distant recipient species such as fish or poultry and not mammals.

One possible way to design primers from coding gene sequences for mapping over radiation hybrid panels could be to design a sequencing primer within the coding region of the gene and to sequence across porcine genomic DNA. This genomic DNA sequence can then be aligned with the coding region from the EST sequence and the exonic region identified. Where short exonic regions are present, then it is likely that the sequence obtained will extend into an intronic region. If this is seen to be the case, then either a marker for RH mapping can be designed directly from the intron region in the sequence read obtained or where more sequence is required, additional sequencing primers can be designed and a larger region of the genomic DNA sequenced.

By comparing gene maps across several species it is possible to investigate potential evolutionary chromosomal rearrangement histories and relationships in gene order between species. The main aim of the human genome project was to complete a high-resolution gene map of the whole genome. This can then to be used to provide a detailed resource for unravelling gene action controlling for example disease susceptibility, reproduction and ageing and also to aid in the unravelling of the evolutionary history of mankind (O'Brien et al., 1993).

The genetics of the laboratory mouse (*Mus musculus domesticus*) has been extensively studied due to its ease of breeding and the use of inbred strains. Although the mouse is a suitable model for molecular studies, it is a poor model for physiological studies (Bishop, 2000). The rat is much more appropriate for these types of studies and indeed a detailed radiation hybrid map of the rat has been produced (Watanabe et al., 1999). In addition, low-resolution genetic maps are now available for more than twenty-eight mammalian species, including several domestic species (O'Brien and Graves, 1991). It is therefore possible to use this information to look in depth at the pattern of chromosomal evolution across mammalian species.

Rodents represent an early branch among placental mammals (Li et al., 1990). Janke *et al.* (1994) estimated that rodents diverged from ferungulates (carnivores and artiodactyls) and primates around 100 million years ago and ferungulates diverged from primates around 80 million years ago. This explains the higher level of conservation of synteny seen between species such as pig and cattle and human than with mice (Johansson et al., 1995). Although since the completion of the draft mouse genome sequence in December 2002, it is now suggested that primates and mice are more closely related than previously believed and that the divergence of the two lineages occurred 75 million years ago (Waterston et al., 2002). The Eutherian placental mammals diverged from the marsupial metatherian mammals 150 million years ago and their common ancestor diverged from the egg laying protherian mammals 170 million years ago. Mammals, birds and reptiles shared a common ancestor around 350 million years ago and interestingly reptiles are believed to have diverged from fish around 400 million years ago (Bishop, 2000). Indeed the level of

conservation of chromosomal synteny between more evolutionary distant species such as chickens from eutherian mammals is very low (Archibald, 1998). Burt *et al.* (1999) predict that one would expect 84 to 600 conserved segments in a comparison between the chicken and human genomes. Interestingly Waddington (2000) estimated that the number of conserved segments between chicken and mouse are almost double that of between human and chickens.

Figure 4-5 demonstrates how a large region of HSA4 is conserved on a single chromosome in mice, sheep, pigs and cattle. Interestingly Burt (2002) describes conserved synteny between a large region of human chromosome 4, including the genes *IL2, KIT, IL8, SPP1, ALB* and *MADH1* and the majority of chicken chromosome 4. The order of homologous genes mapped in sheep and cattle is almost identical (Echard *et al.*, 1994 and (Hediger et al., 1991). More genes have been located in cattle than sheep and therefore the gene locations can be transferred to sheep maps relatively easily. For example Cockett *et al.* (1994) used bovine microsatellite markers to map the callipyge muscle hypertrophy gene in sheep. The causative mutation of this gene has recently been identified (Freking et al., 2002).

Montgomery *et al.* (1995b) describe how, due to the inversion of gene order, the region containing the gene *SPP1* is much closer to the q arm telomere of mouse chromosome 5 and pig chromosome 8 than on human chromosome 4 or sheep and cattle chromosomes 6. Montgomery *et al.* (1995b) also discuss that the region around the Booroola gene (*BMPR1B*), which is linked to *SPP1* in sheep, remains intact in humans, sheep and cattle and these species have single or twin births. Whereas the mouse and pig have much larger litter sizes and there is a breakpoint and inversion of the region around the homologue of *BMPR1B*.

Figure 4-5 confirms that the inversion of this region to the telomere of the q arm has only occurred in pigs and mice. Although the genes *SPP1* and in particular *BMPR1B* do not map as close to the telomere of SSC8 as was previously thought. In addition the order of genes on sheep and cattle chromosomes 6 is very different to that of mouse chromosome 5, pig chromosome 8 and human chromosome 5; where the

order between these three species is relatively much more similar. It is difficult to reason why this particular change in gene order might be connected to the number of offspring a species has.

As mentioned previously Souza *et al.* (2001) and Wilson *et al.* (2001) mapped the Booroola fecundity gene on sheep chromosome 6 around 4 or 5 cM from *SPP1*. Interestingly the physical map of HSA4 shows that *SPP1* is around 10 Mb (equivalent to around 10 cM) away from *BMPR1B* and on the radiation hybrid map of porcine chromosome 8, *BMPR1B* was mapped around 120 cR (equivalent to approximately 9 cM) away from *SPP1*. *BMPR1B* has in fact been mapped using radiation hybrid data to mouse chromosome 3 (MGI Accession ID: 1933683 http://www.informatics.jax.org/searches/reference.cgi?69850). It would therefore appear from Figure 4-5, that the gene order between HSA4 and MMU5 is the same (apart from the far p arm telomeric region being translocated to the telomere of the q arm) and that the breakpoint has occurred between the genes *SPP1* and *BMPR1B*, such that *SPP1* maps to MMU5 and *BMPR1B* maps to MMU3.

The most detailed comparative mapping study of HSA4 and SSC8 carried out to date was performed by Jiang *et al.* (2002a). The results were published at the same time as my radiation hybrid map of SSC8 was completed. Not all of the same genes have been mapped in both studies and it is therefore possible to investigate the results of the two comparative maps built simultaneously and to compare and contrast homologous gene orders. This will give a better understanding of the patterns of gene order conservation between HSA4 and SSC8.

Jiang *et al.* (2002a) used a similar methodology to the experiment described in this chapter, utilising a pig-mouse radiation hybrid panel created with 5 Krads of irradiation. They successfully mapped 46 homologues of human genes to the porcine genome. They aligned the RH map built with a sequence map of HSA4 as well as cytogenetic pig and human maps. They describe that HSA4 is split into nine conserved segments, seven with SSC8, one with SSC15 and one with SSC17. The order of the genes within four out of the seven segments on SSC8 is reversed with

respect to HSA4. Figure 4-6A shows an adapted version of the comparative map built by Jiang *et al.* (2002a) and alongside in Figure 4-6B is shown the equivalent map built using the data from this chapter. Although the general pattern of conservation across the two studies is very similar, Jiang *et al.* (2002a) had mapped several more genes to SSC8 and could therefore define the regions of inversion and translocation with more precision, than was possible for this experiment.

It can be seen that the order of the genes *EGF* and *IL2* are reversed between the maps of SSC8 in Figure 4-6A and Figure 4-6B. If the order of these two markers should indeed have been reversed on my map, then the "pink region" could have been extended. When higher resolution gene maps of SSC8 have been built and the order of genes on HSA4 is determined with more certainty, the definition of the conserved segments and gene order can be ascertained with more certainty. In addition Waddington (2000) describes a mathematical model for estimating the extent of conserved segments, which allows one to predict undiscovered segments using data from known blocks of conserved synteny between two species.

The position of the gene *CPE* on the radiation hybrid map of SSC8 relative to its position on HSA4 (Figure 4-2) was initially a surprise, despite the confidence of mapping it next to the microsatellite marker *S0376* being so high (LOD score 18.2 – see Figure 3-6). In contrast, Cargill *et al.* (1998) had mapped this gene next to the microsatellite marker S0086, which maps about 400 cR from *S0376*, between the genes *SCYB6* and *EGF* (Figure 4-2). However the comparative map built by Jiang *et al.* (2002a) confirmed that this translocation of a small region around CPE is most likely to be genuine. They had mapped the gene *KLHL2* to the same region on SSC8 (the red segment on Figure 4-6A). The detailed gene map of HSA4 obtained from Ensembl maps *KLHL2* at around 163 Mb and *CPE* nearby at around 172 Mb and it is therefore likely that these two genes will be co-located in the same region. In order to be certain of the definite position of *CPE* on SSC8, further genes will need to be added to the RH map in this region.

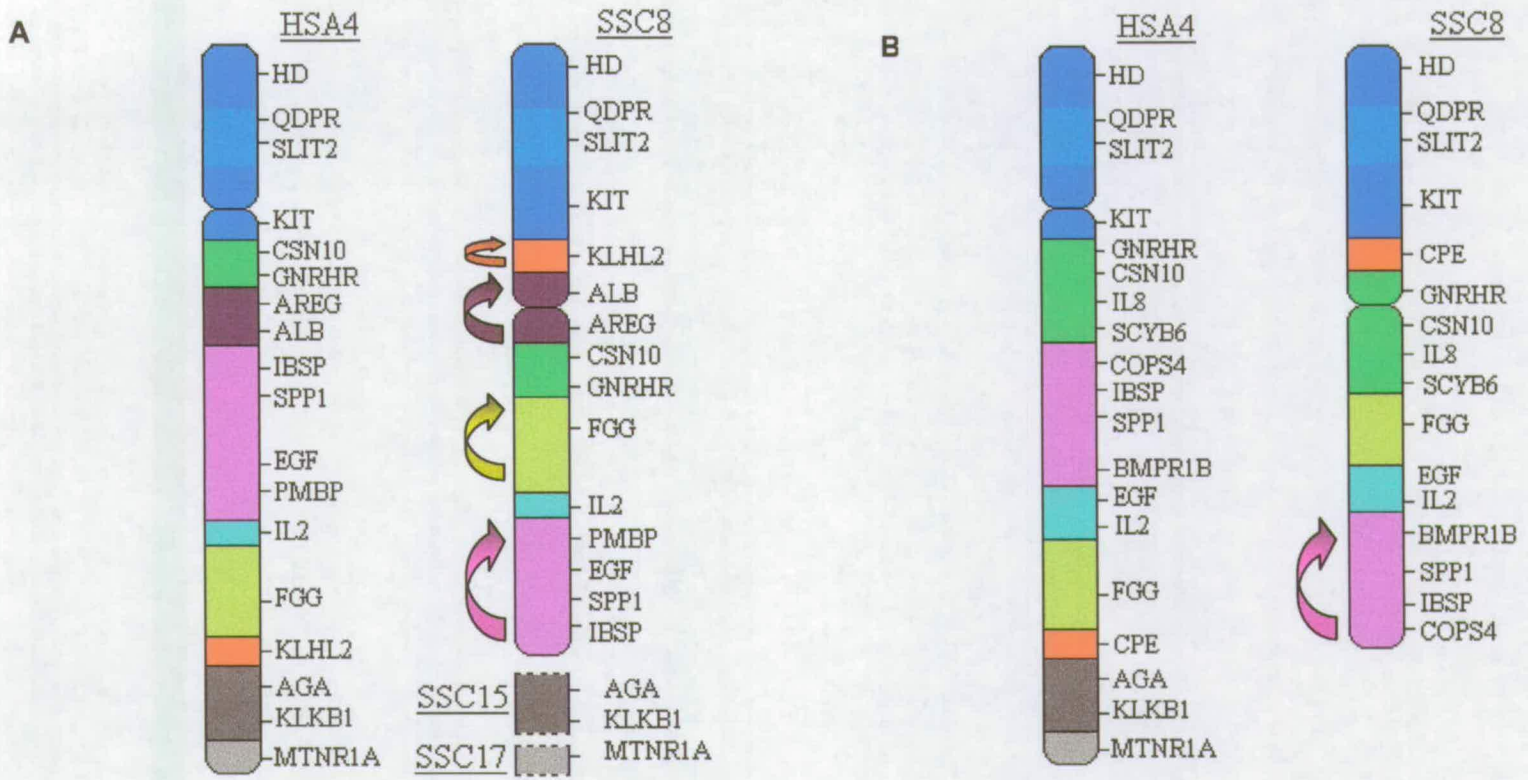


Figure 4-6 (A) Adaptation of the comparative map of human chromosome 4 (HSA4) with the porcine genome, built by Jiang *et* al, 2002. (B) equivalent comparative map of human chromosome 4 (HSA4) and pig chromosome 8 (SSC8) built from the findings of this experiment. Coloured segments are homologous across chromosomes and an arrow indicates an inversion in gene order within that segment.

These detailed comparative maps of HSA4 versus SSC8 built by myself and Jiang *et al.* (2002a) are a valuable tool for the positional cloning of causative genes underlying the economically important QTL mapped to SSC8.

The teat number QTL (age group one animals) mapped around the *SLIT2* gene (Figure 2-6B) and as mentioned in section 2.4 this gene is unlikely to be a putative candidate gene for this trait. It can be seen from Figure 4-2 that this region maps much closer to the p arm telomere on the RH map than was predicted from the linkage map and that the only gene marker that was mapped near to *SLIT2* was *MAN2B2*. *MAN2B2* is not a positional candidate gene for teat number as it maps next to the marker *SW2611*, which can clearly be seen from Figure 2-6B to lie far outside the QTL region. Campbell and Rohrer (2000) investigated *MAN2B2* (alpha mannosidase) as a potential candidate gene for the ovulation rate QTL mapped to SSC8 pter by Rohrer *et al.* (1999). This gene plays an important role in determining the glycosylation status of follicle stimulating hormone and luteinising hormone. However they found no evidence for an association with the genotypes at the polymorphisms investigated in the gene sequence and ovulation rate. They state that the gene is still a plausible candidate in the control of ovulation rate and they are investigating associations with additional polymorphisms.

In addition, the two QTL model for teat number (age group one animals) highlighted a second teat number QTL around the *AREG* gene. The additional genes that were mapped to this centromeric region were *CSN10*, *CSN1*, *AMBN*, *SCYB6* and *IL8*; *GNRHR* and *STE* were already mapped on the linkage map of SSC8. Milk casein can be separated into three components, *CSN10* (kappa-casein), *CSN1* (alpha-casein) and *CSN2* (beta-casein), all of which are closely linked on human chromosome 4. *AMBN* (ameloblastin) is associated with mineralised tissues and in rats it has been shown to be expressed only in teeth, specifically to the ameloblast in the incisors (Krebsbach et al., 1996). *AMBN* a strong candidate gene for amylogenesis imperfecta, a disorder of teeth (MacDougall et al., 1997) and is therefore not a physiological candidate for ovulation rate or teat number.

SCYB6 (small inducible cytokine subfamily B, member 6) and *IL8* (interleukin 8) are both granulocyte chemotactic proteins (Proost et al., 1993). Chemokines regulate cell trafficking of various types of leukocytes and play a fundamental role in the development, homeostasis, and function of the immune system (Fong et al., 1998). It is therefore also unlikely that these genes are candidates for the reproductive traits of interest.

STE (estrogen-preferring sulfotransferase) removes the sulfate group from the precursor estrone sulfate to form active estrogens. The placental and brain estrogen sulfotransferase gene has been mapped to human chromosome 16 (Bernier et al., 1994) and the liver estrogen sulfotransferase cDNA to chromosome 4 (Her et al., 1995). Finally *GNRHR* (gonadotrophin releasing hormone receptor) has already been identified as a strong candidate gene for the ovulation rate QTL mapped around the centromere of SSC8 by myself (chapter 2) and by Wilkie *et al.* (1999) and Braunschweig *et al.* (2001).

Of most interest is that the detailed comparative maps now built between human chromosome 4 and porcine chromosome 8, have allowed the relative position of the prenatal survival QTL around the telomere of SSC8 q arm to be located on human chromosome 4. The QTL maps between the markers *SW61* and *S0178* near the end of the q arm of SSC8; previously only two genes had been mapped to this region (see Figure 2-6C). It can be seen from Figure 4-2 and Figure 4-3 that the equivalent region on HSA4 is from around 82 Mb to around 95 Mb (It is uncertain whether the genes *BMP3, HNRPD* and *HNRPDL* would map to the telomere of SSC8 q arm). Figure 4-7 shows a close up of this specific region taken from Figure 4-2. It can be seen that five genes have now been mapped to this region in pigs and it is discussed below whether there is a role of these genes in controlling prenatal survival and litter size. In contrast up to twenty-three human genes have been mapped in this homologous region and one could be very confident from the comparative maps built, that all the genes in this small region would also map to the QTL region on SSC8. Therefore if further candidate genes are required to be investigated in this region, then the role of these human genes can be researched.

BMPR1B (the pig homologue of the Booroola fecundity gene) appears to map just outside this QTL region (Figure 4-2). A copy of the FecB mutation increases litter size in sheep by one extra lamb. This is achieved through an increased ovulation rate caused by advanced maturation of ovulatory follicles (Souza *et al.*, 2001 and (Wilson et al., 2001). The litter size QTL effects on SSC8 appear to be attributable to improvements in embryo survival and no QTL for ovulation rate were identified in this region. It is therefore possible to exclude *BMPR1B* as a candidate gene for the prenatal survival/litter size QTL. Recently Kim *et al* mapped *BMPR1B* on the USDA-MARC linkage map of SSC8 and they found it to be linked to the marker *SW790* (Kim et al., 2003). It can be seen from the alignment of my RH map with the USDA-MARC linkage map (Figure 3-9) that on the RH map *BMPR1B* was linked to *SW194* and on the USDA-MARC linkage map *SW194* is linked to *SW790*; therefore it appears that *BMPR1B* has been mapped to the same location on both maps. Kim *et al.* (2003) also state that *BMPR1B* maps outside the region of the uterine capacity QTL identified on the q arm of SSC8 by Rohrer *et al.* (1999).

SPARCL1 (SPARC-like protein 1, high endothelial venule protein, hevin) is an extracellular matrix glycoprotein with a function in cell adhesion and proliferation; it is believed to have anti-adhesive properties similar to those of SPARC (Girard and Springer, 1995). It is expressed in several tissues including ovary and lower levels in the placenta (Girard and Springer, 1995). Bertani *et al.* (2002) used differential display PCR to investigate differences in gene expression in the anterior pituitary of a group of control pigs and the Nebraska line selected for increased ovulation rate and embryo survival over several generations. Interestingly they found that *SPARCL1* was differentially expressed between the two groups of pigs and they mention that this gene maps within their age at puberty QTL, which also maps to the telomere of the q arm of SSC8 (Cassady et al., 2001).

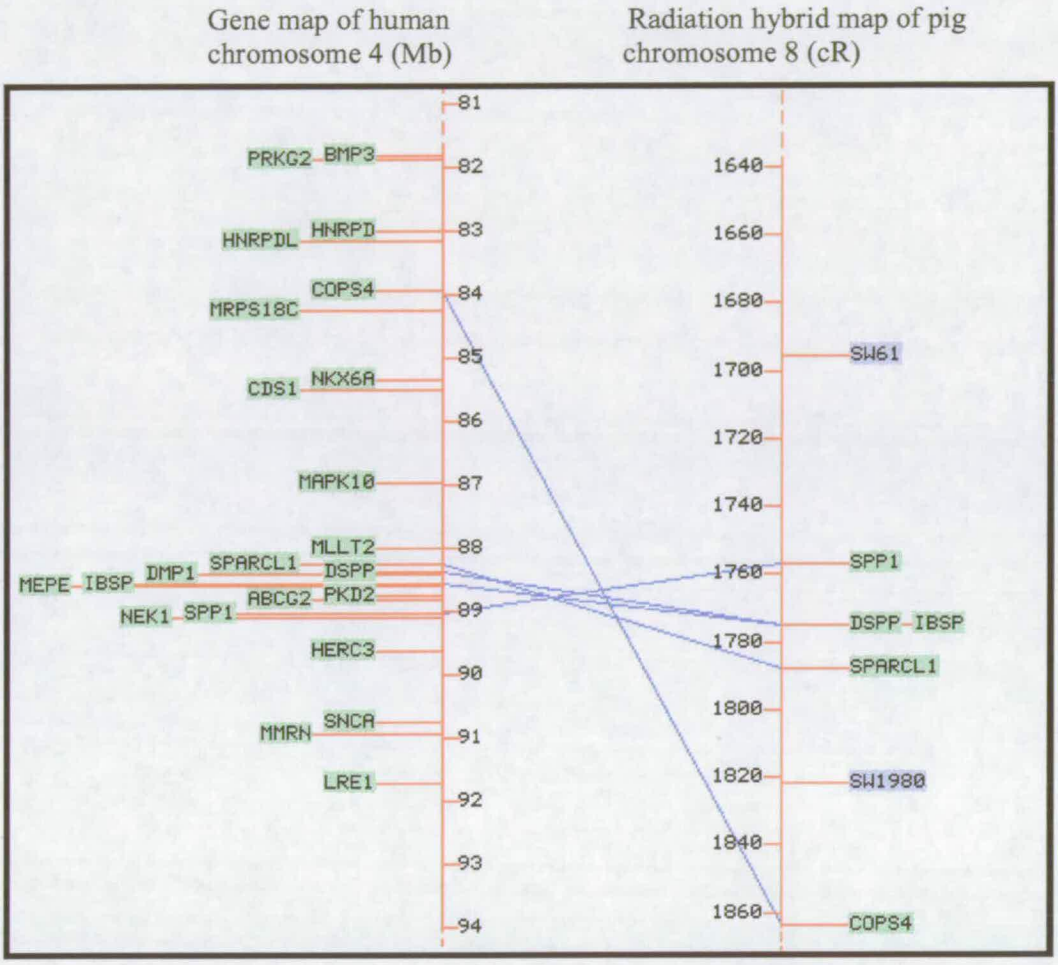


Figure 4-7 Close up of the region of the prenatal survival QTL on SSC8 and the homologous region on HSA4 (taken from Figure 4-2). Gene associated markers are highlighted in green and microsatellite markers are highlighted in blue.

DSPP (Dentin Sialophosphoprotein) is a highly acidic protein and is the major non-collagenous component of dentin. It is solely expressed by the ectomesenchymal derived odontoblast cells of the tooth. Takagi and Sasaki (1988) suggested that a deficiency of this protein is a causative factor in dentinogenesis imperfecta and this gene is therefore clearly not a candidate for a role in controlling embryo survival levels and subsequent litter size in pigs. *COPS4* is a human homolog of an arabidopsis gene known as COP9 signalosome homolog subunit 4 and would therefore also not appear to be a physiological candidate.

IBSP (integrin-binding sialoprotein), like SPP1, is an acidic glycoprotein that undergoes extensive posttranslational modifications. It constitutes approximately 12 % of the noncollagenous proteins in human bone and is synthesized by skeletal-associated cell types, including hypertrophic chondrocytes, osteoblasts, osteocytes, and osteoclasts. Interestingly the only extra skeletal site of its synthesis reported is the trophoblast (Kartsogiannis et al., 1999).

PMBP or *PGRMC2* (progesterone receptor membrane component 2) maps to human chromosome 4 and has been shown to be preferentially expressed in the placenta (Gerdes et al., 1998). The gene maps between *FGG* and *IL2* on human chromosome 4 and although it was not possible to map PGRMC2 to the RH map of SSC8, from the comparative mapping information this gene would be predicted to map at around 1300 cR on Figure 4-2 (near marker S0225). Indeed Jiang *et al.* (2002a) mapped the gene between the markers *IL2* and *EGF*. Therefore although *PMBP* (*PGRMC2*) may be predicted to be a candidate gene, it does not map within the prenatal survival/ litter size QTL regions.

In humans the genes *IBSP*, *SPP1*, *DSPP*, *AMBN* and *BMP3* are all associated with mineralised tissues and all map close together on chromosome 4 (Figure 4-2). Due to the inversion event between porcine chromosome 8 and human chromosome 4, these genes do not map together in pigs (Figure 4-2). *SPP1*, *IBSP* and *DSPP* map to the telomere of the q arm and *AMBN* to the centromere. In humans *BMP3* maps between *AREG* and *DSPP*, at the location of the inversion breakpoint. I had been keen to map this gene to the porcine RH map of SSC8 in order to refine the location of where the breakpoint occurs and to determine whether *BMP3* does indeed map to the prenatal survival QTL region around SSC8qter. However unfortunately the PCR screen had been unsuccessful. *BMP3* (bone morphogenetic protein 3) is a member of the transforming growth factor-beta supergene family and has been shown to induce bone formation (Tabas et al., 1991). It is therefore still not known whether it might be a candidate gene for prenatal survival on physiological or positional grounds.

In conclusion there are two known potential physiological candidate genes now mapped within the prenatal survival QTL regions on SSC8, *SPARCL1* and *SPP1*. As mentioned in chapter 2, *SPP1* remains the strongest physiological and positional candidate gene for the control of prenatal survival rates. The sequence, structure and role of this gene in the early embryo implantation period were therefore investigated in detail (chapter 5). *SPARCL1* is currently being investigated by the Nebraska group (Bertani et al., 2002) and one could confidently predict that the homologues of the human genes shown in Figure 4-7 will also map within the prenatal survival QTL region. Therefore future work could involve the investigation of other potential physiological candidates from this selection of genes.

Chapter Five

5. DETERMINATION OF SPP1 GENE STRUCTURE AND SNP IDENTIFICATION

5.1. Introduction

Evidence suggests that there is a QTL around the telomere of the q arm of porcine chromosome 8, controlling litter size and the related trait of prenatal survival. The beneficial alleles at this locus appear to be from the Meishan breed (Figure 2-6C and Table 2-5). The gene, known as secreted phosphoprotein 1 (*SPP1*) or osteopontin (*OPN*), maps within the 95 % confidence interval for this QTL. Indeed it lies directly under the highest estimated F value on the QTL plot, this being the most likely position of the QTL (Figure 2-6C).

On its own this finding does not provide sufficient evidence to justify investigating this positional candidate gene. The confidence interval for the QTL is large and thus there are many positional candidate genes. However, there are other lines of evidence that indicate that *SPP1* merits further examination as a candidate gene. First, van der Steen *et al.* (1997) demonstrated a positive association between two alleles at a microsatellite locus (5' of the coding region of *SPP1*) and an increase in litter size in a Meishan 50 % synthetic line. Following this, a polymorphic marker (*SPP1-1*) based on the same microsatellite repeat was typed and included in the QTL analysis (chapter 2).

More recently, Korwin-Kossakowska *et al.* (2002) reported a positive association between the presence of a SINE (a mammalian-wide interspersed repeat element) in the *SPP1* gene and higher litter size in a commercial Polish pig line. Individuals that were homozygous for the presence of the SINE had an average litter size (total number born) of 11.98 piglets. Individuals homozygous for the absence of the SINE had 11.08 piglets and heterozygous animals 10.92 piglets. The difference between these genotype classes was significant ($P < 0.05$) and demonstrated that the additive effect was positive (+0.45 piglets) and the dominance effect negative (-0.61 piglets). A negative dominance effect indicates negative overdominance, where the performance of the heterozygotes is inferior to both classes of homozygotes. The

same trend in the genetic effects was observed for the QTL for prenatal survival level and litter size (Table 2-5).

These reports of association between litter size and *SPP1* marker genotypes, however, may simply indicate the presence of a gene influencing little size, which is closely linked to but distinct from *SPP1*. Similarly, in the QTL analysis (chapter 2) it may be *SPP1* or a closely linked gene that is influencing the trait.

The justification for considering *SPP1* further mainly comes from complementary lines of research on the function of this gene. Physiological studies demonstrate strong evidence for a role of *SPP1* in embryo implantation and placentation in many species, including pigs (Fazleabas *et al.*, 1997; Garlow *et al.*, 2002; Johnson *et al.*, 1999b; Johnson *et al.*, 1999a; Johnson *et al.*, 2000; Omigbodun *et al.*, 1997; Waterhouse *et al.*, 1992 and{Johnson, 2001 160 /id}. These studies show that during the peri-implantation period glands of the endometrium secrete histotroph, which contains several endometrial proteins including *SPP1*. This histotroph nourishes and sustains the conceptus and *SPP1* has been shown to promote remodelling of the trophectoderm, adhesion of the trophectoderm with the endometrium and the formation of the placenta (Johnson *et al.*, 1999b; Johnson *et al.*, 2000 and {Garlow, 2002 344 /id}. It has been reported that the significant differences in prenatal survival levels between the Meishan breed and US and European commercial breeds can be mainly attributed to the marked reduction in peri-implantation conceptus loss seen in the Meishan breed (Ford, 1997; Vonnahme *et al.*, 2002; Wilson, 2002{Ford, 1997 271 /id} and (Wilson et al., 1999). Even when the uterus size and the ovulation rate of the two breeds has been observed to be similar, the Meishan breed farrows three to five more viable piglets per litter (Lee *et al.*, 1995 and Haley and Lee, 1993).

Therefore there is positional and physiological evidence to indicate *SPP1* as a strong candidate gene controlling differences in prenatal survival level and subsequent litter size. The evidence suggests that investigation of the different strategies utilised by

the two breeds during the peri-implantation period and the potential role of *SPP1* during this period would be worthwhile.

The specific aim of this study was to sequence a copy of the entire gene (including around 2.5 kb 5' of exon 1) from Large White and from Meishan breed origin. Previously, only the coding and promoter regions of the gene had been sequenced in pigs (Wrana *et al.*, 1989 and (Zhang et al., 1992a). The two gene sequences could then be compared with the purpose of identifying single nucleotide polymorphisms (SNPs) and other sequence variants in the exons and introns of the gene and also in the 5' promoter region. The level of sequence conservation in the different regions of the gene could then be investigated to determine whether any variants are found to lie within regulatory regions of the gene. Another aim was to determine whether those SNPs that result in a change in the amino acid sequence consequently change the secondary structure of the protein formed.

Therefore this study has been the first to deduce the entire sequence of the *SPP1* gene and promoter regions in a European breed of pig (Large White). It has also allowed the direct comparison of the gene sequence with that of the breed of interest with respect to improved litter size, the Meishan breed.

5.2. Materials and methods

(Protocols for all solutions mentioned are detailed in appendix I)

5.2.1. Southern hybridisation of PigE BAC library with target DNA probe

5.2.1.1. Preparation of PigE BAC filters

Southern hybridisation experiments involve the annealing of denatured single stranded radiolabelled DNA probes to complementary DNA fragments immobilised on membranes. The idea of immobilising a hybridising molecule on a solid support was first proposed by Denhardt (1966) and the transfer of desired DNA molecules to the membrane developed by Southern (1975).

The Roslin PigE BAC (porcine bacterial artificial chromosome) library provides approximately five-fold coverage of the porcine genome and consists of 102,912 clones, with each individual clone containing an insert of double stranded DNA, from a copy of a single chromosome, of around 150 kb (Anderson et al., 2000). The DNA used to create the library was obtained from peripheral blood lymphocytes of an F1 Large White / Meishan boar (for more detail http://www.hgmp.mrc.ac.uk/geneservice/reagents/products/descriptions/pig_BAC.shtml). The library is commercially available as high-density gridded filters for screening and the individual clones are stored in LB agar + 7.5 % glycerol within single wells of 268 384-well microtiter plates (filters and clones are available from the UK HGMP resource centre). The library can be screened by hybridisation with the 102,912 clones gridded in duplicate in a 4x4 array onto six 22.2 x 22.2 cm Hybond N nylon membranes (Amersham Pharmacia Biotech, Amersham Life Science Ltd, UK).

The set of six filters had been stored wet wrapped in Saran wrap at -20° C. These six membranes were first pre-hybridised in aqueous prehybridisation/hybridisation (APH) solution. This solution contains reagents that block non-specific DNA

binding sites on the membrane surface thereby reducing unwanted background hybridisation.

By alternating with 4 layers of nylon membrane (Amersham Pharmacia Biotech, Amersham Life Science Ltd, UK) the first three filters were carefully unwrapped from the Saran wrap and using tweezers placed into a half full tray of APH solution, with the label of the filter in the top left corner of the tray.

Again using tweezers the first three filters were loosely rolled up and placed into a glass cylinder (Techne, Cambridge, UK) and 20 ml APH solution added. This was repeated for the last three filters and both cylinders placed in a revolving hybridiser HB-1D oven (Techne, Cambridge, UK) at 65° C for 30 minutes.

5.2.1.2. Radioactive labelling of DNA probe with [α32P]dCTP

The method of priming random hexanucleotides of all possible sequences to a denatured DNA probe template, was first developed by Feinberg and Vogelstein (1983 and 1984). A complementary DNA strand is synthesised by Klenow polymerase using dATP, dTTP, dGTP and [α^{32}P]dCTP by priming to the 3' hydroxy termini of the random oligonucleotides. This produces a radiolabelled probe of the desired DNA template sequence, where all the cytosine bases in the sequence are labelled.

Template DNA (25 ng) was added to double distilled water to a final volume of 11 µl into a 1.5 ml eppendorf tube. This DNA sample was then denatured at 95° C for 10 minutes on PHC-2 thermocycler (Techne, Cambridge, UK).

The sample was immediately placed on ice and 4 µl High Prime labelling mixture (Roche Diagnostics, Mannheim, Germany) added. This mixture contains random 6-mer oligonucleotides, 0.125 mM dATP, dGTP and dTTP, 5x stabilized reaction buffer in 50 % glycerol and 1 Uμl^{-1} Klenow polymerase. Taking the necessary precautions for use of radioactivity, 5 µl ^{32}P dCTP was added to the diluted DNA sample. This was then incubated for 10 minutes in a 37° C water bath, with a lead

cover over the sample tube to shield the radioactive emissions. The reaction was stopped and the radioactively labelled DNA denatured by adding 5 µl 2M NaOH.

5.2.1.3. Hybridisation and auto radiography

The labelled DNA was then diluted with 500 µl APH solution and 250 µl pipetted into each of the hybridisation cylinders containing the filters and the 20 ml APH solution. The bottles were mixed and incubated in the hybridiser HB-1D oven at 65° C overnight.

The hybridisation probe mix from both bottles was disposed of and 50 ml of wash 1 added to remove excess probe and the cylinders returned to the hybridisation oven for 15 minutes. This wash was repeated, 100 ml of wash 2 was then added to each cylinder. Wash 2 has a lower salt concentration than wash 1 and removes any non-specific binding. The cylinders were returned to the oven for 30 minutes and again wash 2 was repeated.

The membranes were removed from the cylinders and the level of radioactivity on the filters confirmed to be less than 5 counts per second. Each individual filter was carefully blotted and wrapped in Saran wrap. It was then placed between intensifying screens in an autoradiography cassette with the label of the filter in the top left corner and secured in place.

In the dark room, an X-ray film (AGFA-Curix Blue) was placed on top of the filters inside each of the 6 cassettes. The cassettes were then placed in a -80° C freezer overnight. The X-ray films were developed using an automatic developer (X-Ograph Compact x2).

The positions of the positive clones for the DNA probe on the filters were determined first by identifying the filter number on which the duplicate positive signals are located. Then the position of the label on the filter was used to identify the panel in which the clones are located. The plate within that panel was determined by the orientation of the two positive signals within the 4x4 array on the

panel. The co-ordinates on the 384 well plates were determined by the location of the array on the panel by referring to the row (A-P) and the column (1-24). The location of the positive clones is given as the plate number - row and column.

5.2.2. Obtain single colonies of clones and vector purification

5.2.2.1. Preparation of LB (Luria-Bertani) plates

It can be seen from Figure 5-1 that the pBeloBAC 11 vector used to clone the genomic BAC library contains a chloramphenicol antibiotic resistance gene. It was therefore necessary to grow colonies from the positive BAC clones on agar containing the same antibiotic. 500 ml of LB bottom agar was melted and 250 µl of 25 mg/ml chloramphenicol added. Individual sterile petri plates were poured and left to set for around 60 minutes. All equipment was flamed in order to reduce contamination.

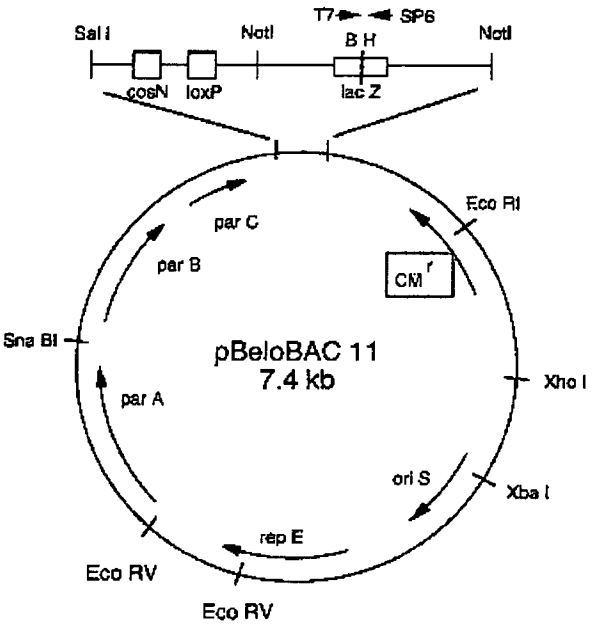


Figure 5-1 pBeloBAC 11 vector used to clone porcine genomic PigE BAC library

By selecting the appropriate wells from the PigE BAC library plates containing the positive clones, the cells in that well were mixed, colonies picked and streaked onto the agar plates. This was repeated for each of the positive clones identified. These plates were then incubated in a 37° C oven overnight to allow colony growth.

5.2.2.2. *Preparation of LB Liquid broth*

37.5 µl of 25 $mgml^{-1}$ chloramphenicol was added to 150 ml of liquid LB broth and 5 ml volumes aliquoted into universal tubes. Single colonies for each positive clone were picked into these tubes using a sterile pipette tip and incubated overnight at 37° C in an incubator shaker at 300 rpm (New Brunswick Scientific, Edison, NI, USA). It was necessary to pick a single colony to confirm that only the clone containing the desired DNA insert was selected, as there is always a small risk of contamination from neighbouring wells containing different clones.

5.2.2.3. *Preparation of glycerol stocks*

500 µl of liquid broth from each colony grown was dispensed into a small universal tube, spun at 1400 x *g* for 10 minutes using an IEC centra-8R centrifuge (International Equipment Company, USA), the supernatant removed and 250 µl of LB with 25 % glycerol added. These were stored at -80° C.

5.2.2.4. *Vector DNA purification*

The QIAprep spin miniprep kit (Qiagen Ltd, Crawley, UK) was used to isolate the plasmid containing the desired DNA insert from the bacterial lysate. The liquid broth colonies were spun at 1400 x *g* (IEC centra-8R centrifuge) for 10 minutes, the supernatant removed and the pellets re-suspended in 250 µl buffer P1. The re-suspended pellets were transferred to 1.5 ml eppendorf tubes and 250 µl buffer P2 added. These tubes were inverted 4-6 times, until the solution turned clear. 350 µl buffer N3 was added and the tubes inverted 4-6 times. They were spun at 12,000 x *g* (Micro centaur, MSE) for 8 minutes and the supernatant transferred into a new eppendorf tube.

200 µl ProCipitate™ (Liochem) was added to the supernatant and spun for 5 minutes at 12,000 x *g* (Micro centaur, MSE). The supernatant was transferred to a new

eppendorf tube and 600 µl isopropanol added. These were stored on ice for 30 minutes and then spun at 12,000 x *g* (Micro centaur, MSE) for 10-15 minutes until a pellet was formed. The supernatant was removed and the pellet air-dried for a few minutes. 50 µl of TE was added and the tubes vortex mixed.

Eluted DNA (2 µl) was run with loading dye on a 25 ml 0.8 % agarose gel (SeaKem® ME agarose, RMC BioProducts, Maine, USA) in 1x EB buffer stained with 10 µg of ethidium bromide (Sigma-Aldrich Company Ltd, Dorset, UK) to check that a sufficient concentration of plasmid DNA had been purified for each clone.

5.2.3. Primer design

All primers were designed using the web based program Primer3 (http://www-genome.wi.mit.edu/genome_software/other/primer3.html) (Rozen and Skaletsky, 2000). The sequence of the region of DNA for amplification was imported into the program and left (forward) and right (reverse) primers were designed accordingly. The primers were then ordered from MWG Biotech AG, Ebersberg, Germany and diluted in $T_{10}E_1$ to a stock concentration of 100 pmolμl^{-1}.

5.2.4. Purification of target DNA PCR products

5.2.4.1. Purification of PCR products from gel extraction

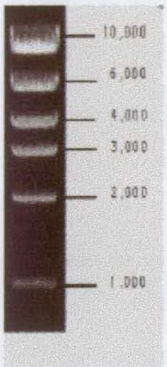
PCR products were purified from the amplification products using QIAquick® Gel Extraction Kit (Qiagen Ltd, Crawley, UK) following the protocol for use with a microcentrifuge. All steps of the protocol were followed, the DNA was eluted in 30 µl double distilled water and the columns were allowed to stand for 1 minute.

5.2.4.2. Determination of gel purified DNA concentration

The concentration of the gel purified DNA was determined by running 1 µl of the DNA alongside 2 µl and 4 µl of High DNA Mass™ ladder (Invitrogen Life Technologies, Paisley, UK) through 0.8 % multipurpose (MP) agarose gel (Roche Diagnostics, Mannheim, Germany) stained with 40 µg ethidium bromide. This

ladder is composed of an equimolar mixture of six DNA fragments. Figure 5-2 shows the size of the ladder bands and the concentration of DNA in each band.

Size in bp



Fragment size (bp)	2µl of ladder	4µl of ladder
10,000	100 ng	200 ng
6,000	60 ng	120 ng
4,000	40 ng	80 ng
3,000	30 ng	60 ng
2,000	20 ng	40 ng
1,000	10 ng	20 ng

Figure 5-2 Size and concentration of DNA fragments within the High DNA Mass™ ladder.

5.2.5. DNA Sequencing (ABI 373 sequencer)

The ABI PRISM™ dye terminator cycle sequencing ready reaction kit (Applied Biosystems, Warrington, UK) was used for sequencing sample preparation. This ready reaction mix contains AmpliTaq® DNA polymerase FS; Tris-HCl, pH 9.0; $MgCl_2$; thermal stable pyrophosphatase buffer; dNTPs and dye terminator dNTPs.

Sequencing reactions were carried out to a total volume of 10 µl; 4 µl ready reaction mix was added to 1 µl of DNA (gel purified PCR product), 1 µl of sequencing primer (3.3 $pmol\mu l^{-1}$) and 4 µl of double distilled water into 200 µl eppendorf tubes. Where a higher concentration of DNA was required part of the water was substituted with additional DNA. The subsequent PCR reactions consisted of 10 seconds denaturation at 96° C, 5 sec at 55° C for primer annealing and 4 min at 60° C for primer extension and dye termination. The reactions were carried out for 35 cycles in a Perkin-Elmer Gene Amp-9700 thermal cycler.

5.2.5.1. *Acrylamide gel preparation*

The glass plates, spacers and comb were washed with Alconox® (Aldrich chemical co., Milwaukee, USA), rinsed with distilled water and air-dried. The dried plates were laid horizontally in the correct orientation and aligned with the spacers (0.4 mm).

The gel consisted of 40 ml Ultra Pure SequaGel® XR monomer solution (National Diagnostics, Hull, UK), 10 ml Ultra Pure SequaGel® complete buffer reagent (National Diagnostics, Hull, UK) and 350 µl 10 % Ammonium Persulphate (APS). Using a 50 ml syringe attached to a 0.8 µm filter (Millipore) the gel was carefully poured between the plates ensuring no bubbles were formed. The comb was inserted, the plates clamped and the gel left to polymerise for around one hour. After polymerisation the clamps and comb were removed and the outsides of the plates were washed with distilled water and dried. The gel plate was set-up as per the instructions for the Applied Biosystems (ABI) model 373A automated DNA sequencing system. The gel was pre-run for half an hour to warm up the gel and prepare it for loading. The 373 sequencer uses slab-gel electrophoresis, the gel is formed by cross linking polyacrylamide and bis-acrylamide and the degree of crosslinking affects the porosity of the gel and therefore the resolution of separation of the DNA fragments.

5.2.5.2. *Precipitation of samples and gel loading*

The PCR dye-terminated products were precipitated with 35 µl 95 % ethanol at 4° C and the tubes placed on ice for 10 minutes. The samples were then centrifuged at 13,000 x *g* (8000 series, Centurion) for 20 minutes. The supernatant was carefully removed and the tubes air-dried for 30 minutes. The pellets were then resuspended in 6 µl of formamide loading buffer (Amersham Pharmacia Biotech, Amersham Life Science Ltd, UK) and denatured for 5 minutes at 94° C. The denatured samples were kept on ice prior to loading into the sequencing gel and 3 µl of each sample was loaded into alternate lanes of the gel. The remaining 3 µl were stored at -20° C for use in a repeat sequence run if required. The gel was then pre-run for 5 minutes, stopped and then the remainder of the samples loaded. This loading procedure was

followed in order to prevent spill over between adjacent lanes. The gel was run for 14 hours at 30 mA in 1 x TBE buffer.

5.2.6. Analysis of sequence quality using Chromas

The sequence traces produced for each sample run was analysed using the program Chromas (available from Technelysium Pty Ltd, (http://www.technelysium.com.au/chromas.html). The overall quality of the sequence and the length of the sequence read were assessed.

5.2.7. Sequence assembly and SNP identification

Sequence assembly and analysis was carried out using the Staden Package version 2001.0 (Staden, 1996).

5.2.7.1. Preparation of sequence reads using Pregap4

Each individual sequence read was saved in the format of an ABI file and these files were imported into the program Pregap4. This program converts the trace file format (*.ABI) into standard chromatogram format (*.SCF), creates an experiment file (*.exp) for each trace and estimates confidence values for the accuracy of each base and consequently highlights any poor quality sequence.

The option of "interactive clipping" within the program Trev (Bonfield et al., 2002) was then used to manually check each read and confirm or extend the areas of poor quality sequence. Pregap4 then creates a file containing all of the quality sequence reads, with the areas of poor quality masked.

5.2.7.2. Assembly of sequence reads using Gap4

Normal shotgun assembly allowing a maximum mismatch level of 4 % was used to assemble the sequence reads into contigs. Within this program it is possible to look at the sequences directly and therefore investigate any uncertainties. The orientation of the contigs was checked and the option of "find internal joins" was used to join

the contigs together. The consensus sequence from all the reads was then exported as a FASTA file.

5.2.7.3. *Alignment of sequences to identify putative variants*

The FASTA sequence files were aligned using a multiple sequence alignment program (Corpet, 1988), in order to identify variation between them. The alignment was then viewed and annotated within the program GeneDoc (Nicholas and Nicholas, 1997).

5.2.7.4. *BLAST searching for sequence homology*

The BLAST search program (http://www.ncbi.nlm.nih.gov/BLAST/) was developed by Altschul *et al.* (1997) and involves the use of local pairwise alignments between either nucleotide or amino acid sequences. An alignment score is produced by searching a "word size" of every 11 bases along the sequence of interest against all sequences present in the various databases mentioned. This therefore allows gaps to be introduced when matching the two sequences and an overall level of alignment ascertained. The output also reports a percentage identity between two sequences, however it is best to consider the alignment score, which is measured in "BITS", as the most reliable prediction of how accurate the match is. This BIT score is assigned an E value depending on how significant the match is. In order to avoid finding chance matches, a threshold of significance is produced from a model of random sequences. E values which are less than $1e^{-20}$ can be considered to be significant. As a rough rule of thumb this represents an alignment with a bit score greater than 80.

5.3. Results

5.3.1. Identification of PigE BAC clones containing SPP1 gene sequence

In order to ascertain the complete sequence of the *SPP1* gene, it was necessary to obtain genomic DNA containing the gene sequence. Therefore the Roslin PigE BAC library (Anderson et al., 2000) was screened with a DNA probe from an *SPP1* cDNA clone to identify single clones containing part or the whole of this gene sequence. The aim was to sequence a copy of the gene from both the Large White and Meishan breeds of pigs and to investigate whether variation in the gene contributed to the differences seen in embryo survival and litter size between breeds. The library was prepared from an F1 purebred Large White x purebred Meishan boar and it was therefore possible to obtain copies of the gene from alternative chromosomes from the two breed origins.

The full length cDNA clone of the porcine secreted phosphoprotein-1 (*SPP1*) gene (EMBL accession no. X16575) described by Wrana *et al.* (1989) was available in our laboratory. The clone (~1.5 kb) was gel purified from the pT7T3-10B vector. Figure 5-3 shows the structure of the vector containing the cDNA clone. The restriction enzyme *Bam*H1 was used to cut out the insert and the resulting product gel purified. The DNA fragment used to screen the Roslin PigE BAC library therefore contained a small amount of vector either side of the complete *SPP1* cDNA.

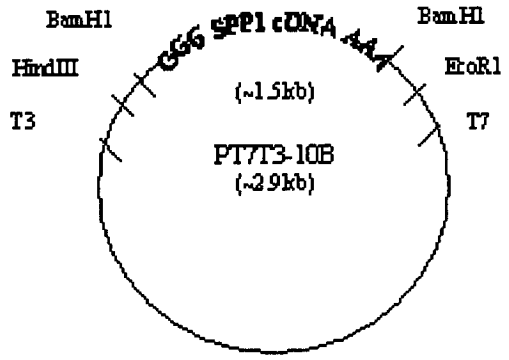


Figure 5-3 Structure of PT7T3-10B vector containing *SPP1* cDNA insert

A radioactive probe was prepared for the cDNA sequence, to screen the Roslin PigE BAC library for clones containing the *SPP1* gene. By running 1µl of the cDNA insert along side High DNA Mass™ Ladder (Invitrogen Life technologies) on 1 % MP agarose gel stained with 40 µg ethidium bromide, it was possible to estimate the concentration of the insert DNA to be only around 2.5 ng μl^{-1} (Figure 5-4). The ideal concentration of template DNA for labelling is a minimum of 25 ng, therefore the remaining 11 µl of undiluted stock DNA was used to make the probe.

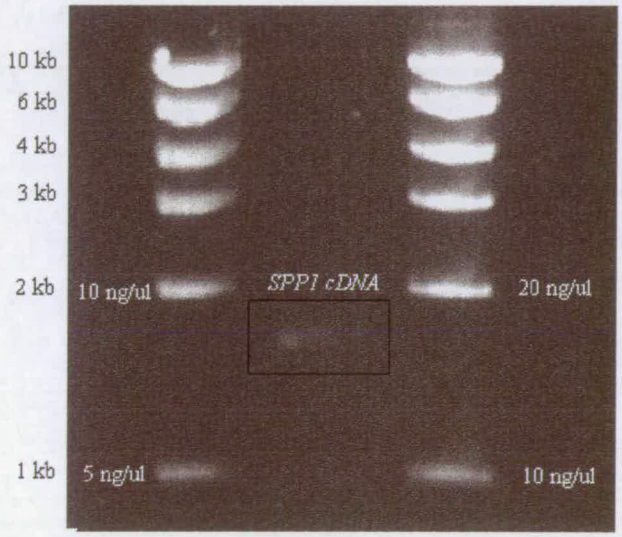


Figure 5-4 Determination of the concentration of the porcine cDNA insert

Seven positive BAC clones were identified to contain sequence from the *SPP1* gene: 44c21, 120d12, 123m22, 150a11, 157-g11/12, 173-a13/14 and 201n13. The location of the clones on plates 157 and 173 was uncertain because the position of the positive label on the filter screen appeared to overlap two wells on the 384 well plates. It was therefore decided that both clones that the probe appeared to bind to, should be tested for the presence of the *SPP1* gene.

5.3.2. Confirm presence of entire SPP1 gene from Large White and Meishan breeds in positive BAC clones

It was necessary to identify which of the positive BAC clones contained sequence for the entire gene and also to locate a copy of the gene inherited from the purebred Large White parent and a copy from the Meishan.

5.3.2.1. Screen positive clones to confirm presence of SPP1 gene

Each of the positive clones was tested to confirm the presence of the *SPP1* gene using PCR. Known primers from the 5' region of intron 1 were used, forward 5'-CCGCTGATGGTTGCTGTC-3' and reverse 5'-ATCTGCCCTGGATCTGAC-3'. A 9 µl reaction mix was made into 0.5 ml eppendorf tubes. The mix consisted of 10 pmol of each primer (MWG Biotech AG, Ebersberg, Germany), 2.5 mM $MgCl_2$ in 1x PCR buffer II (Applied Biosystems, Warrington, UK), 2.5 mM of dTTP, dCTP, dGTP and dATP (Amersham Pharmacia Biotech inc, Little Chalfont, UK) and 1 U AmpliTaq gold (Applied Biosystems, Warrington, UK) and the edge of a single colony was picked into the PCR mix. The location of the colony picked was marked on the plate. A couple of drops of oil were added to seal the PCR mix. The PCR reactions were carried out on TRIO-thermoblock PCR machine (Biometra) using the following program:

Denaturation cycle:	94° C 12 min
10x cycles:	94° C 20 sec
	52° C 30 sec
	72° C 30 sec
25x cycles:	94° C 20 sec
	58° C 30 sec
	72° C 30 sec
Extension cycle:	72° C 5 min

To confirm which of the clones amplified the specific 150 bp PCR product from intron 1 of the *SPP1* gene, 9 µl of each of the PCR products were separated by electrophoresis through a 25 ml 4 % MP agarose gel in 1x EB buffer stained with 10 µg ethidium bromide.

BAC clones 44c21 and 201n13 were the only ones out of the possible nine positive clones to contain the region of *SPP1* intron 1 tested for. Consequently the remainder of the single colony chosen for each of these two clones was picked into LB broth. Glycerol stocks were made to ensure that additional stock DNA of the specific colony picked would be available if required. The BAC DNA was then purified using the QIAprep spin miniprep kit (Qiagen Ltd, Crawley, UK) and confirmed by gel electrophoresis to be of sufficient concentration.

5.3.2.2. Identify clones containing entire SPP1 gene

Primers were designed that flanked or lay within six of the seven exons of the gene and within the promoter region 5' of exon 1 (Table 5-1). The sequence information for primer design was obtained from published porcine sequences (GenBank accession numbers M84121, X16575 and AJ237667).

Table 5-1 Primers designed around exons and promoter region of *SPP1*

Region of *SPP1* gene	Forward primer	Reverse primer	Product size (base pairs)
Promoter	5'-AACCATAGTGAATCCTGCGG-3'	5'-GCTCTGAGCCCATTTGAAAC-3'	550
Exon 1	5'-GATGTCTGGTGCAGCCTTTA-3'	5'-TCCCCGTGAAATGAAACAGT-3'	192
Exon 2	5'-TGTGGTGGCTTGAAAAGATG-3'	5'-TGTACTCACTGGAAGGGCAGA-3'	149
Exon 4	5'-GCTTTCCAACAAATACACAGATG-3'	5'-CTGTGGCGCTAGGAAAGTCT-3'	82
Exon 5	5'-GAGGAAACGGACGACTTCAA-3'	5'-GTCAGCGTGATCAGCTTCCT-3'	144
Exon 6	5'-AGCAACCGACGTCACTCC-3'	5'-ATACATGAGCCTGCCGATTC-3'	194
Exon 7	5'-CATCCCTGGCTGTTCATTTA-3'	5'-GGCTGACTCGTCTCCTGACT-3'	198

PCR was carried out on the purified DNA from the BAC clones 201n13 and 44c21 using all of these primer pairs. A 10 µl reaction mix was made for each clone and primer pair. This PCR mix consisted of 5 pmol of each primer (MWG Biotech AG, Ebersberg, Germany), 1.5 mM $MgCl_2$ in 1x PCR buffer 1 (Enzyme Technologies Ltd, Cambridge, UK), 2.5 mM of each dTTP, dCTP, dGTP and dATP (Amersham Pharmacia Biotech inc, Little Chalfont, UK), 0.25 U super taq (Enzyme technologies Ltd, Cambridge, UK) and 1 µl of purified DNA diluted 1 in 100 with double distilled water.

The GeneAmp PCR system 9700 (Perkin Elmer™, Foster City, CA, USA) was used with a touchdown PCR program:

Denaturation cycle:	94° C 5 min
7x cycles:	94° C 20 sec
	62° C 20 sec* Decrease 1° C per cycle to 56° C
	72° C 30 sec
25x cycles:	94° C 20 sec
	56° C 20 sec
	72° C 30 sec
Extension:	72° C 5 min

In order to determine which of the BAC clones contained the whole of the *SPP1* gene, 1 µl of PCR product from each of the regions of the gene was run with 100 bp DNA ladder (Invitrogen Life Technologies) through a 25 ml 4 % MP agarose gel (Roche Diagnostics, Mannheim, Germany) in 1x EB buffer stained with 10 µg ethidium bromide. It can be seen from Figure 5-5 that the porcine DNA insert from both BAC clones 201n13 and 44c21 appeared to contain the entire *SPP1* gene. The region amplified by the primers within exon 5 was seen to be much larger than expected and those in exon 7 also amplified a region of just over 300 bp as well as the expected amplicon of 198 bp. However the results of the gel in Figure 5-5 provided sufficient evidence to prove that the whole of *SPP1* from the promoter region to exon 7 is present in the two BAC clones. It was later realised that exon 5

was smaller than predicted and that the reverse primer believed to be located in exon 5 actually lay at the 3' end of intron V and consequently the region amplified by these two primers was in fact over 1 kb and not the expected size of 144 bp (see position of the primer SPPEx5R on Figure 5-12).

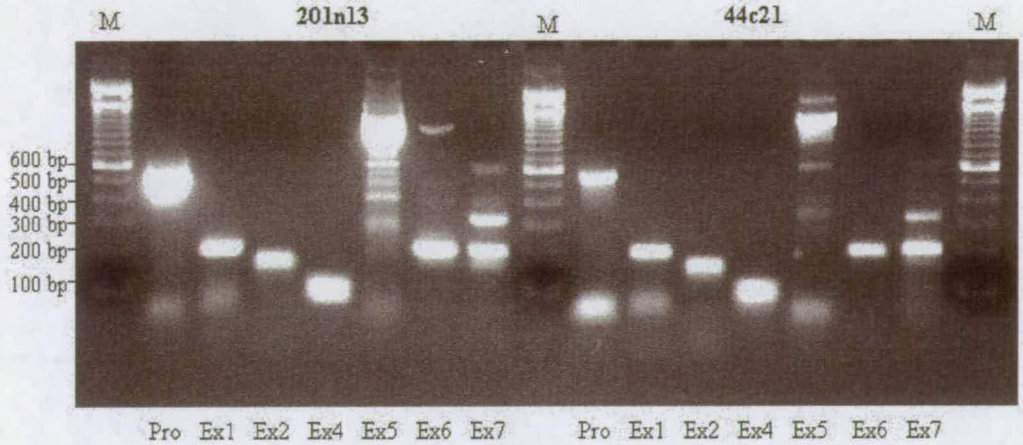


Figure 5-5 PCR products of BAC clones 201n13 and 44c21 across seven regions of *SPP1* gene. (M=100 bp ladder, Pro = promoter region (550 bp), Ex1 = exon 1 region (192 bp), Ex2 = exon 2 region (149 bp), Ex4 = exon 4 region (82 bp), Ex5 = exon 5 region (144 bp), Ex6 = exon 6 region (194 bp) and Ex7 = exon 7 region (198 bp).

5.3.2.3. Distinguish copies of the gene from Meishan and Large White

A SNP had been identified previously in a Meishan x Large White crossbred population, at the 5' end of intron 1 (position 2770 of *SPP1* (appendix II)), which resulted in a base change from guanine to thymine (Carole Sargent, personal communication). Where base G was present, this allele cut with the restriction enzyme *Mnl*I. It was necessary to determine whether this polymorphism exhibited breed specific alleles.

Genomic DNA was used from 30 F0 individuals from the three populations outlined in section 2.2.1. and from the F1 boar used to create the BAC library and his purebred Meishan father and Large White mother. The PCR mix and reaction conditions are the same as described in section 5.3.2.2 and the primers used were

those flanking exon 1 (see Table 5-1 (forward = 5'-GATGTCTGGTGCAGCCTTTA-3' and reverse = 5'-TCCCCGTGAAATGAAACAGT-3' and a product size of 192 bp)). In this experiment 50 ng genomic DNA was added into the PCR mix and the GeneAmp PCR system 9700 (Perkin Elmer™, Foster City, CA, USA) was used. To check that the PCR had been successful, 1 µl of PCR product (192 bp) from a few of the samples were run through a 25 ml 4 % MP agarose gel stained with 10 µg ethidium bromide (see Figure 5-6).

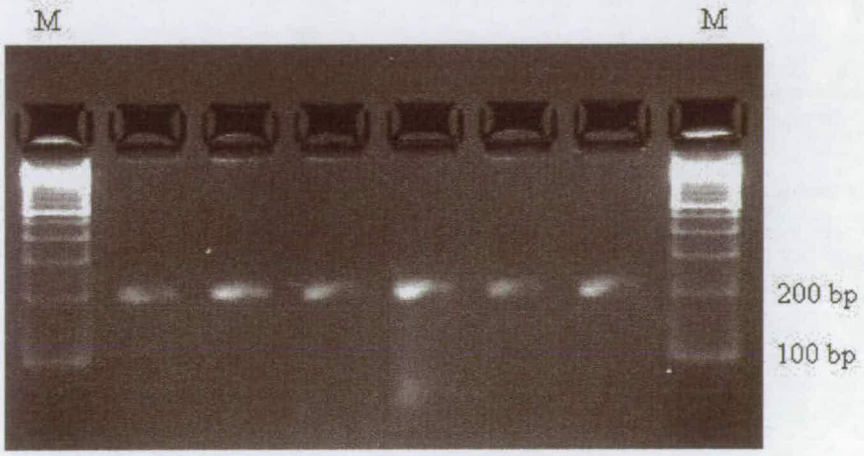


Figure 5-6 Confirmation that PCR products of 192 bp had been successfully amplified for six of the DNA samples. (M= 100 bp DNA ladder).

The PCR products (10 µl) were then digested with 5 U of *Mnl*I, 1x Buffer 2 and 0.01 % BSA (New England Biolabs) at 37° C for two hours. The recognition sequence of *Mnl*I is CCTC. The amplicon for individuals with base G at the SNP, will be digested by *Mnl*I to give five fragments (109 bp, 27 bp, 26 bp, 23 bp and 7 bp) and the amplicon with base T will be digested to give four fragments (132 bp, 27 bp, 26 bp and 7 bp). It is therefore possible to distinguish the 109 bp product from the 132 bp product by running the samples on a high percentage agarose gel alongside a 100 bp DNA ladder. The restriction fragments (2 µl) were run through a 4 % MP agarose gel stained with 40 µg ethidium bromide. The PCR and digest was carried out twice, in order to confirm that the fragments produced were not just a result of partial digestion (Figure 5-7).

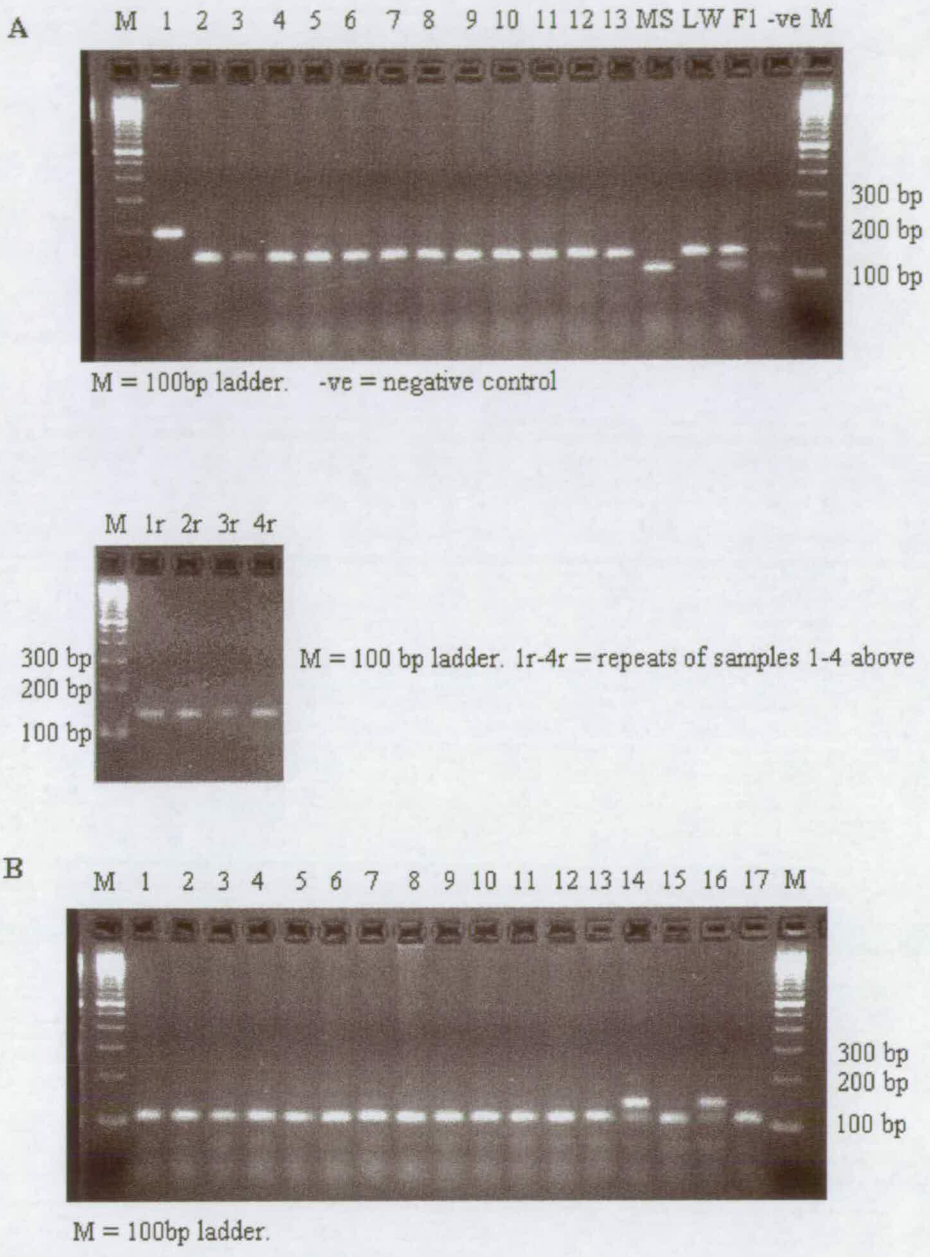


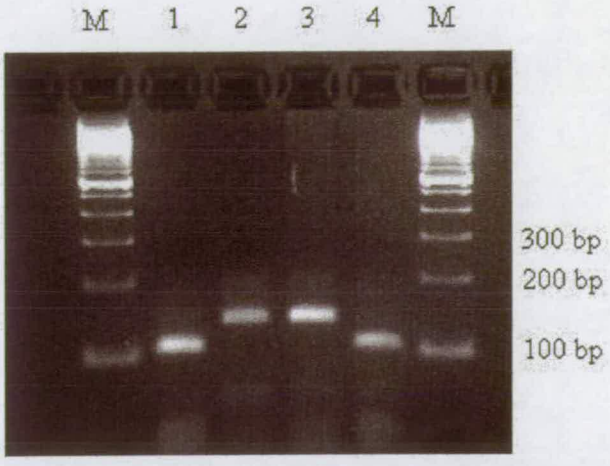
Figure 5-7 *Mnl*I digest of PCR products incorporating a SNP at the 5' end of intron 1 of *SPP1*. Where base G is present at the SNP, the 192 bp PCR product will be digested to give five fragments (109 bp, 27 bp, 26 bp, 23 bp and 7 bp) and where base T is present the digest results in four fragments (132 bp, 27 bp, 26 bp and 7 bp). **A.** Results for 13 F0 Large White individuals (1-13), the F1 boar used to sequence *SPP1* (F1) and his purebred Meishan (MS) and Large White (LW) Parents. The amplicon of individual 1 was undigested, therefore the repeat digest is also shown (1r). **B.** Results for 17 F0 Meishan animals (1-17).

All the Large White individuals were seen to be homozygous for the base *T* allele and all except two of the Meishan individuals (Figure 5-7B, samples 14 and 16) were homozygous for the base *G* allele. Individuals 14 and 16 were both heterozygotes. As noted earlier, the BAC library was prepared from an F1 Large White/Meishan boar and indeed the PCR-RFLP confirmed that this F1 boar was heterozygous for this SNP at the 5' end of intron 1. As expected his Large White dam was homozygous for the *T* allele and his Meishan sire was homozygous for the *G* allele. Therefore, this *MnlI* PCR RFLP test could be used to determine the breed origin of the cloned DNA inserts within the BAC clones shown to contain the *SPP1* gene.

This *MnlI* PCR-RFLP assay was used to test DNA from the two clones identified to contain the entire gene in the same way as described above for genomic DNA. As controls two genomic DNA samples, one from a Meishan individual homozygous for *G* (animal 3 on Figure 5-7B) and one from a Large White individual homozygous for *T* (animal 5 on Figure 5-7A) were also tested. The results are shown in Figure 5-8. BAC clone 44c21 was shown to possess base T at this SNP and therefore appears to contain the Large White copy of the gene and BAC clone 201n13 was shown to possess base G and therefore appears to contain the Meishan copy of the gene.

5.3.3. *Obtain DNA fragments across five regions of the SPP1*

Primers were designed to amplify the whole gene by PCR from both BAC clones, representing the two breeds, across five overlapping regions. Published sequences of the gene were used for this primer design, which included the cDNA sequence, a 2 kb promoter region 5' of exon 1, exons 1, 2, 6 and 7 and introns 1 and 6. This amplification across regions allowed the whole gene to be sequenced. The primer sequences are shown in Table 5-2.



1= 201n13, 2= 44c21, 3= Large White, 4= Meishan

Figure 5-8 *MnlI* digest: 44c21 equivalent to Large White (LW), 201n13 equivalent to Meishan (MS). (M = 100 bp ladder).

Table 5-2 Primer sequences used to amplify 5 regions of *SPP1* gene.

	Primer name	Primer sequence	Product size
Region 1 (Promoter and Exon 1)	SPPaF (forward) SPPE1R (reverse)	5'-GAATTCACTCGTCTTTCCTTTGA-3' 5'-TCCCCGTGAAATGAAACAGT-3'	~2.8 kb
Region 2 (Exons 1, 2 and 3 and introns 1 and 2)	SPPE1F (forward) SPPE3R (reverse)	5'-GATGTCTGGTGCAGCCTTTA-3' 5'-TCCGAGCTGCCAGAATTAGT-3'	~1.5 kb
Region 3 (Exon 2, 3 and 4 and introns 2 and 3)	SPPE2F (forward) SPPE4R (reverse)	5'-TGTGGTGGCTTGAAAAGATG-3' 5'-CTGTGGCGCTAGGAAAGTCT-3'	~3.1 kb
Region 4 (Exon 4, 5 and 6 and introns 4 and 5)	SPPE4F (forward) SPPE6R (reverse)	5'-GCTTTCCAACAAATACACAGATG-3' 5'-ATACATGAGCCTGCCGATTC-3'	~1.8 kb
Region 5 (Exon 6 and 7 and intron 6 and part 3'UTR)	SPPE6F (forward) SPPendR (reverse)	5'-AGCAACCGACGTCACTCC-3' 5'-TTCTCCCACCTCGCTACAAT-3'	~1.4/1.7 kb

The PCR mix and reaction conditions are the same as described in section 5.3.2.2. 40 µl PCR reaction mixes were made for each primer pair. The purified BAC clone DNA was diluted to an optimised concentration to ensure a sufficient concentration of PCR product. The total product from each region of the gene was run through a 100 ml 1 % MP agarose gel in 1x EB buffer alongside a 1 kb plus DNA ladder (Invitrogen Life Technologies) (Figure 5-9). It was then possible to estimate the size of the entire gene and promoter region to be around 10 kb.

Table 5-2 shows the approximate size of each region of the gene amplified. It was noticed that region 5 in the Large White copy of the gene was around 300 bp larger than the Meishan copy. Investigation of the published sequence across exons 6 and 7 and intron 6 (GenBank accession number AJ237667) and the cDNA sequence (GenBank accession number X16575) showed that the region amplified by the primers SPPE6F and SPPendR would be expected to be around 1.7 kb.

The published sequence AJ237667 is from a Landrace breed of pig and Knoll *et al.* (1999) describe how there is variation between individuals for the presence or absence of a block of 305 bp in intron 6, consisting of a 291 bp SINE (PRE-1) sequence and a 14 bp flanking sequence. I concluded that the Meishan copy of the gene probably did not contain the SINE repeat element and the Large White copy did. The sequencing results confirmed this hypothesis.

Finally, the PCR products for the five regions of the gene from both BAC clones were purified using QIAquick® Gel Extraction Kit (Qiagen). The PCR and gel purification was repeated to give enough purified DNA to sequence all five regions of the gene for both breed origins. The purified DNA from all the PCR reactions across each region was combined and the concentration of DNA determined using High DNA Mass™ ladder (Invitrogen Life Technologies) (Figure 5-10). By combining products from different PCR reactions, it reduced the possibility of errors being introduced, where non-complementary bases may have been incorporated into a single DNA copy during the polymerase reaction.

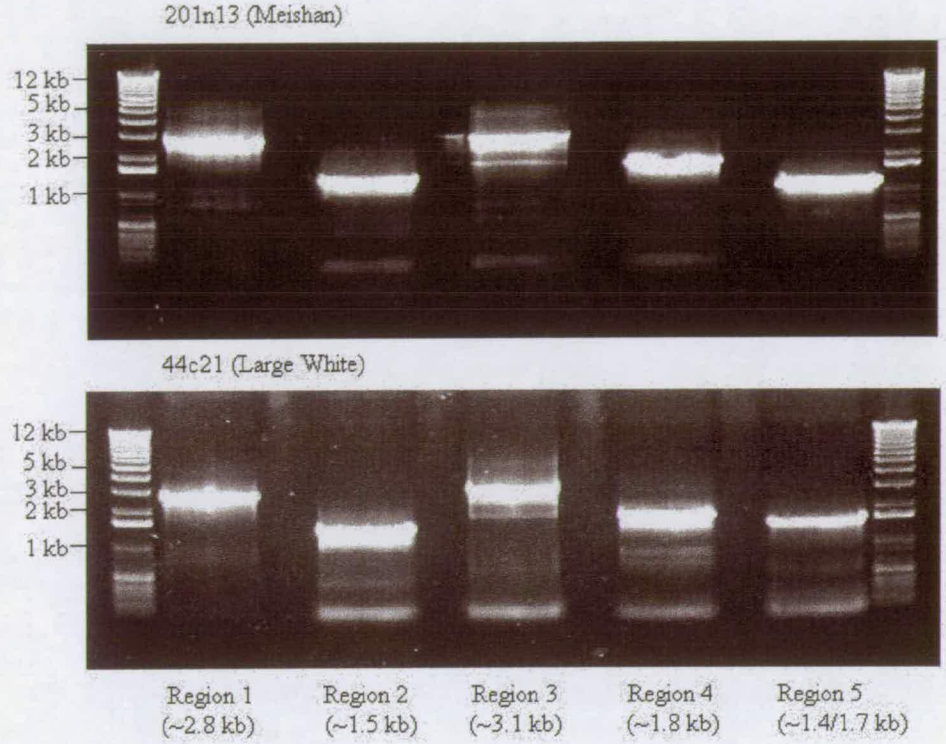


Figure 5-9 PCR products from five regions of Large White and Meishan copies of *SPP1*

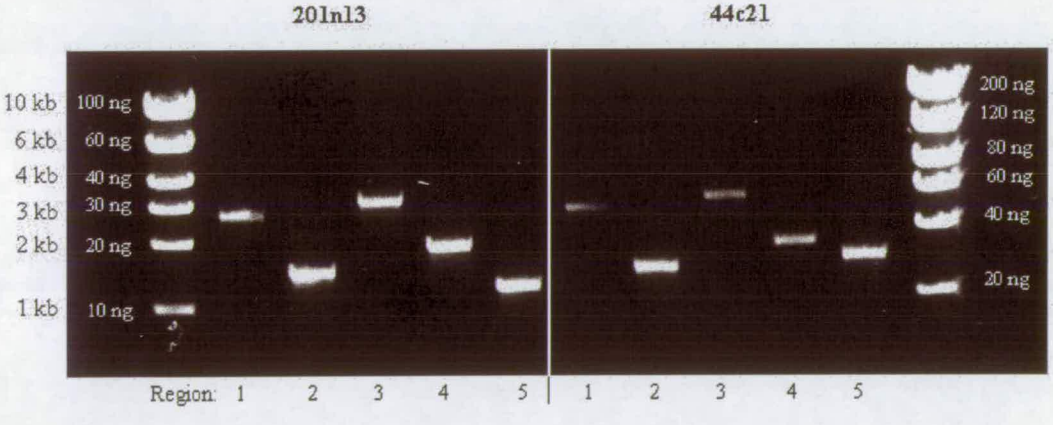


Figure 5-10 Concentration of DNA for sequencing determined using High DNA Mass™ ladder.

For sequencing 20 ngμl^{-1} per kb of DNA is ideally required, however more than 1 μl of DNA can be used in the sequencing reaction mix. It can be seen from Figure 5-10 that regions 1, 3 and 4 from BAC clone 44c21 were more dilute than 20 ngμl^{-1} per kb, therefore a greater volume of DNA was used than for the other regions of the gene in the sequencing reaction.

5.3.4. Primer design and sequencing of the gene

Sequencing of the gene was initially carried out using ABI 373 DNA sequencer. The first sets of sequences were produced from the primers used to amplify the five regions of the gene from the genomic DNA of the Large White and Meishan breeds (see Table 5-2). The sequences produced were then used to design further primers in a technique known as primer walking. Both forward and reverse primers were designed to ensure that the entire gene was sequenced for each breed with at least a two-fold coverage on each strand. In this way, one could be certain that any difference identified between the sequences of the two breeds were not just as a result of a region of poor sequence quality.

A draft sequence of around 7 kb of the gene was produced. However the sequences produced by the ABI 373 DNA sequencer (PE Biosystems) was not of a high enough quality; the maximum length of quality sequence obtained per read was only around 400 bp and a lot of manual editing of the sequence data was required. Therefore all the primers designed and the DNA from the five regions of the gene from both breeds was sent to the DNA Sequencing Facility at the University of Cambridge where an ABI 377 DNA sequencer (Applied Biosystems, Warrington, UK) was used, producing quality reads of 600-700 bp. Figure 5-11 shows a comparison between the same region of sequence from the ABI 373 DNA sequencer (top sequence) and from the ABI 377 DNA sequencer (bottom sequence). It can clearly be seen that the lower sequence is of a much higher quality and more accurate.

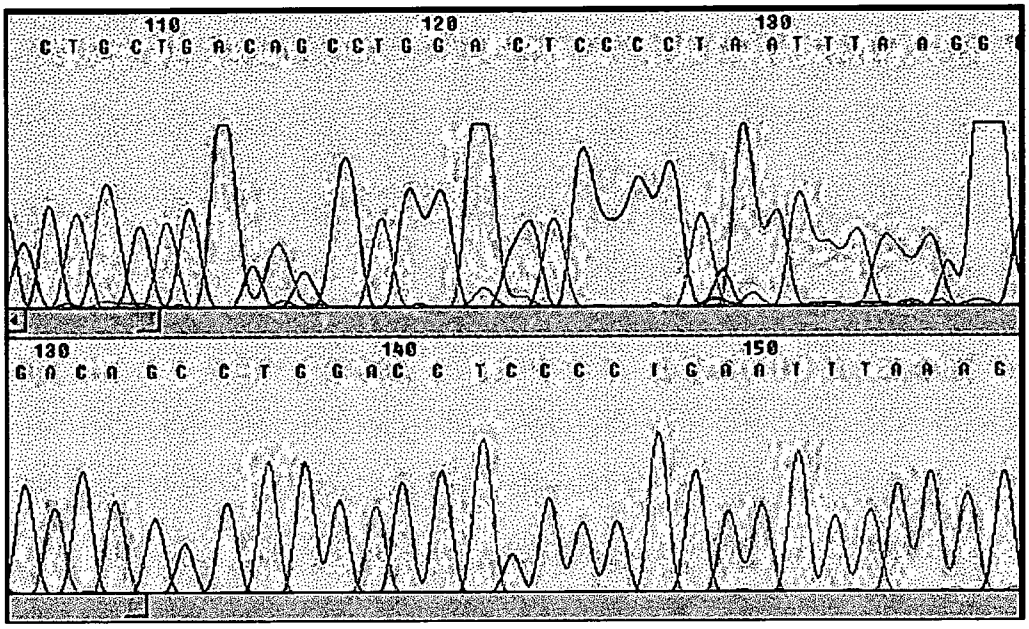


Figure 5-11 A comparison between the same region of sequence from the ABI 373 DNA sequencer (top sequence) and from the 377 sequencer (bottom sequence).

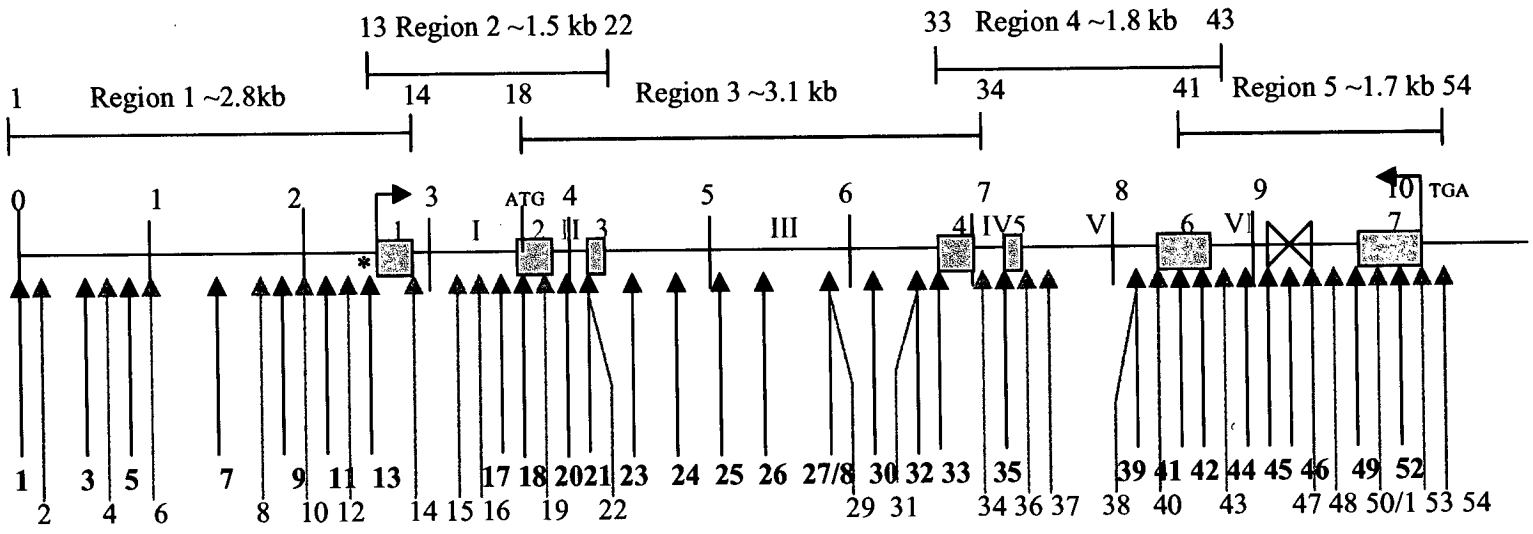
The quality sequence obtained was then used to design primers to complete the entire sequence of the gene. Some regions for example the microsatellite repeat 5' of exon 1 and the SINE repeat element proved difficult to sequence and additional primers were required to be designed around these regions. A total of 54 primers were designed in order to produce a satisfactory quality of sequence over the whole gene (~10kb). The primer sequences are shown in Table 5-3. Figure 5-12 summarises the structure of the *SPP1* gene and the location of all 54 primers.

5.3.5. Sequence assembly and SNP identification

All the sequence reads from the Large White and Meishan copies of the gene were prepared using Pregap4 and then independently assembled using Gap4 within the Staden Package. A single contig was produced from each breed. The two complete sequences were then aligned and the differences between them identified (see Appendix II). Table 5-4 summarises the size of each region of the gene and the number and type of putative variants found within them.

Table 5-3 Summary of 54 primers used to sequence *SPP1* gene and promoter region (F = forward and R = reverse primer).

Primer name	Primer sequence (5'- 3')	Primer name	Primer sequence (5'- 3')
SPPaF	GAATTCACTCGTCTTTCCTTTGA	SPPI3aR	GGGGCTCCTGGAATTAAAAG
SPPaR2	GACCGAGCCATCAAACAACT	SPPI3gF	CACAGAGGCCCGATATGAAG
ProF	AACCATAGTGAATCCTGCGG	SPPI3bR	CAAGTGAGAGTCAGTGGCAGA
SPPaR	CACACATTTTAATAATTCGGATGG	SPPI3eF	TGTGGCTGAATGTCTCCTGA
SPPbF	CGTGAGTCTTACAGTGGAGGTG	SPPI3hF	TGGCCTCATCTGAGAGGAAT
ProR	GCTCTGAGCCCATTTGAAAC	SPPE4F	GCTTTCCAACAAATACACAGATG
SPPcF	TAGCCTGGGAACTTCCATGT	SPPE4R	CTGTGGCGCTAGGAAAGTCT
SPPbR	CCTTGTTCATGGAGGTTTTG	SPPE5F	GAGGAAACGGACGACTTCAA
SPPdF2	CATCACCTCCAAACCATTCA	SPPI5cR	TAACGTAATTTTGCCGGAGT
SPPcR	TCACGTTCAAATCCACGTTC	SPPI5bR	TTCCTGGAGGCAAAATCTTGT
SPPdF	TGGCAGAACTCTTTGTGTGCC	SPPI5aF	CGGAGTGCATTGTCCTGACTG
SPPdR	TGGGTCCCCTAAAAAAGCGGG	SPPI5aR	TCTTCCTTTCGTGTGAGAATCTT
SPPE1F	GATGTCTGGTGCAGCCTTTA	SPPE5R	GTCAGCGTGATCAGCTTCCT
SPPE1R	TCCCCGTGAAATGAAACAGT	SPPE6F	AGCAACCGACGTCACTCC
SPPI1aR	TGTCATGTGGGTATTTTGGTG	SPPE6R	ATACATGAGCCTGCCGATTC
SPPI1bR	GGTGCTACTTAAGTGAAATGAAGC	SPPE6Fc	AAGTTCCGCAGATCCGAAG
SPPI1aF	TGGAAAATAGAGGTGCCCTAA	SPPI6aF	CACATGGCTTTGCCATTTAC
SPPE2F	TGTGGTGGCTTGAAAAGATG	SPPI6bF	ATCACGAGGCAGTTTTTCCA
SPPE2R	TGTACTCACTGGAAGGGCAGA	SPPI6aR	GCGGGTCTCGTCTGTTTTAC
SPPI2aF	CAGAATTCCCCCAAATAAATGA	SPPSINEFb	CTACACCAGAGCCACAGCAA
SPPE3F	AACAGACTAATTCTGGCAGCTC	SPPI6bR	GCTTTCTAACTTTCAGGCCAGT
SPPE3R	TCCGAGCTGCCAGAATTAGT	SPPE7F	CATCCCTGGCTGTTCATTTA
SPPI3aF	TCCAGTAGCACGAAATTCCA	SPPE7R	GGCTGACTCGTCTCCTGACT
SPPI3bF	AAAAAGGAAGTCTATTTATGCTGTTT	SPPE7R2	TCCGTCTCCTCACTTTCCAC
SPPI3cF	TCCAAACTGAGTTAAGAGCACAA	SPPendF	CAAGCTGGTCCCAGACTCTAA
SPPI3fF	GGAACACTGGTTACCTATGAGTGA	SPPendR2	TGATCTCAGAAGACGCACTCTC
SPPI3dF	AAAGAAAAGAAAGAGGGAACTTG	SPPendR	TTCTCCCACCTCGCTACAAT



1 SPPaF	**13 SPPE1F**	**25 SPPI3cF**	38 SPPI5aR	50/51 SPPE7R & R2
2 SPPaR2	14 SPPE1R	**26 SPPI3fF**	39 SPPI5aF	**52 SPPendF**
3 ProF	15 SPPI1aR	**27/28 SPPI3dF+gF**	40 SPPE5R	53 SPPendR2
4 SPPaR	16 SPPI1bR	29 SPPI3aR	**41 SPPE6F**	54 SPPendR
5 SPPbF	**17 SPPI1aF**	**30 SPPI3hF**	**42 SPPE6Fc**	
6 ProR	**18 SPPE2F**	31 SPPI3bR	43 SPPE6R	KEY
7 SPPcF	19 SPPE2R	**32 SPPI3eF**	**44 SPPI6bF**	⋈ Sine element (343bp) not present in MS
8 SPPbR	**20 SPPI2aF**	**33 SPPE4F**	**45 SPPI6aF**	ATG Start translation 1-7 Exons
9 SPPdF2	**21 SPPE3F**	34 SPPE4R	**46 SPPSINEFb**	TGA Stop translation I-VI Introns
10 SPPcR	22 SPPE3R	**35 SPPE5F**	47 SPPI6aR	* Microsatellite repeat (interferes with sequencing)
11 SPPdF	**23 SPPI3aF**	36 SPPI5cR	48 SPPI6bR	↱ Start transcription ↰ End transcription
12 SPPdR	**24 SPPI3bF**	37 SPPI5bR	**49 SPPE7F**	

Figure 5-12 Structure of *SPP1* gene and location of primers used for sequencing. The five regions of PCR products are also shown.

Table 5-4 Number and types of putative sequence variants within each region of the porcine *SPP1* gene (information obtained from figure in appendix II).

Region	Size (bp)	Number of SNPS		Indels	Repeats	Total
		Transition (C-T)	Transversion (C-A, C-G, T-A)			
5' exon 1	2649	3	7	0	1 (microsatellite)	11
Exon 1	84	0	0	0	0	0
Intron 1	1149	0	3	1	0	4
Exon 2	66	0	0	0	0	0
Intron 2	113	0	0	0	0	0
Exon 3	38	0	0	0	0	0
Intron 3	2752	13	21	5	0	39
Exon 4	80	0	0	0	0	0
Intron 4	291	1	3	1	0	5
Exon 5	41	0	0	0	0	0
Intron 5	1075	6	14	2	0	22
Exon 6	308	0	5 (2 at one site)	0	0	5
Intron 6	1015	3	2	1	1 (SINE)	7
Exon 7	386	2	0	0	0	2
3' UTR	57	1	1	0	0	2
TOTAL	10104	29	56	10	2	97

5.3.6. Sequence translation and identification of amino acid variants

The nucleotide sequence from the coding regions of both copies of the gene were translated to amino acid sequence using the translation program within GCG (Womble, 2000). The predicted open reading frame encoded a total of 303 amino acids, in agreement with previously published data (GenBank accession number X16575 and Swissprot accession number P14287 (Wrana et al., 1989)). The sequences were then aligned (Figure 5-13). Three differences were observed in the predicted amino acid sequences of the proteins encoded by the Large White and Meishan forms of the gene. Two of these differences are encoded by SNPs in exon 6. The first is a change at residue 110 from an alanine to a threonine; the region on the nucleotide sequence is at position 8449 to 8451 (appendix II). The second is a change at residue 159 from a valine to an alanine and the nucleotide position is at 8596 to 8598 (appendix II). The third change is encoded by a SNP in exon 7 (position 9919 to 9921 (appendix II)), where a proline is present at residue 262 in the Large White sequence and a serine in the Meishan sequence.

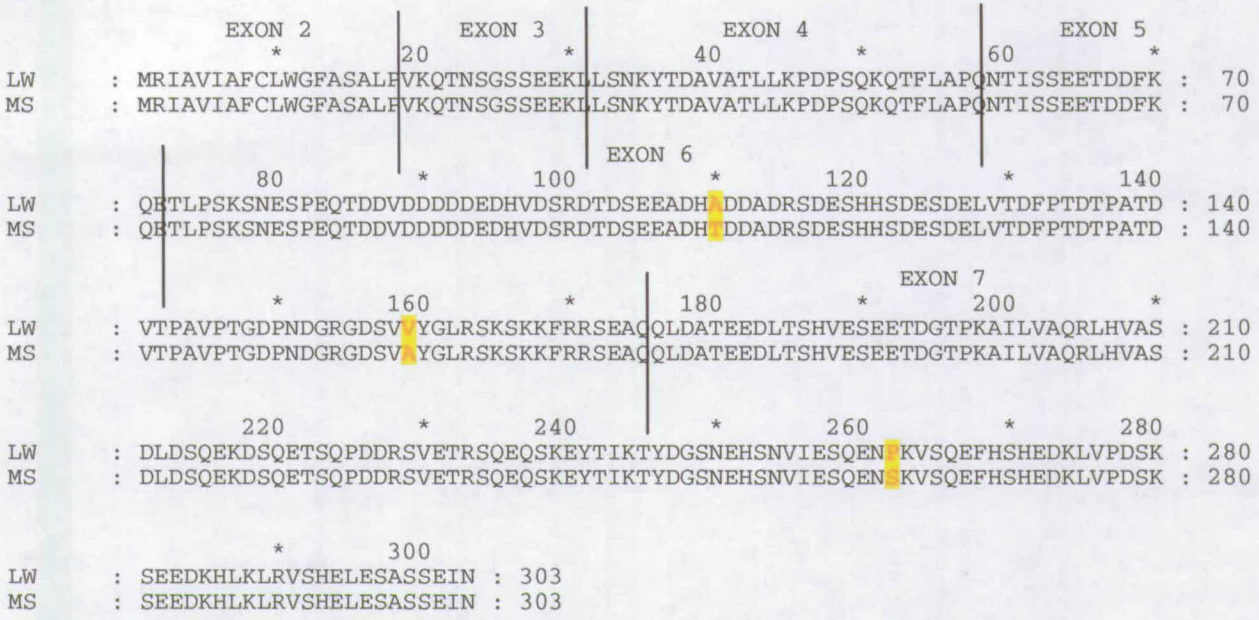


Figure 5-13 Alignment of the predicted amino acid sequence of Meishan (MS) and Large White (LW) SPP1. Putative sequence variants are highlighted. The relationship of the encoded protein sequence to the exons in the gene is shown.

5.3.7. Alignment of nucleotide and published porcine sequences

5.3.7.1. Identification of promoter regulatory regions

The nucleotide sequences for both the Large White and Meishan copies of the gene were aligned with the published porcine sequence of the promoter region of *SPP1* (GenBank accession number M84121) (see Figure 5-14). Zhang *et al.* (1992a) characterised this promoter region by constructing chimeric chloramphenicol acetyltransferase (CAT) constructs containing different regions of the 5' end of the gene, including around 3 kb upstream of exon 1, intron 1 and exons 1 and 2. CAT is a bacterial reporter gene and was contained within an SV40 basic vector lacking any promoters or enhancers. The different regions of the gene were then investigated for their promoter activity. In addition consensus sequences for regulatory motifs were identified using the program DNA Inspector IIe. The consensus sequences and the regions of promoter activity identified are marked on Figure 5-14.

No sequence variants between the Large White and Meishan sequences were observed within these putative regulatory sequences. The region around the microsatellite repeat (highlighted in green), that was shown to possess strong promoter activity, also contained no sequence variants except for the microsatellite repeat itself. Zhang *et al.* (1992a) suggest that there are two promoter regions, the traditional region 5' of the transcription start site in exon 1 and also a possible alternative site in intron 1 and that it is possible that alternative promoters may be used in different tissues resulting in transcripts of varying length. Indeed Saavedra *et al.* (1995) describe that there are two transcripts found in mice, one that contains the untranslated exon 1 and the other starts from the 3' end of intron 1, therefore indicating a splice site within intron 1. Zhang *et al.* (1992a) suggest that that the 3' half of intron 1 from base number +690 on Figure 5-14 contains the most significant promoter activity. There is one SNP and one base insertion difference between the two breeds in this 3' region of intron 1; however the overall level of variation in intron 1 is lower than the other introns in the gene, indicating conservation of this region between the two breeds. A BLAST search of the published pig promoter

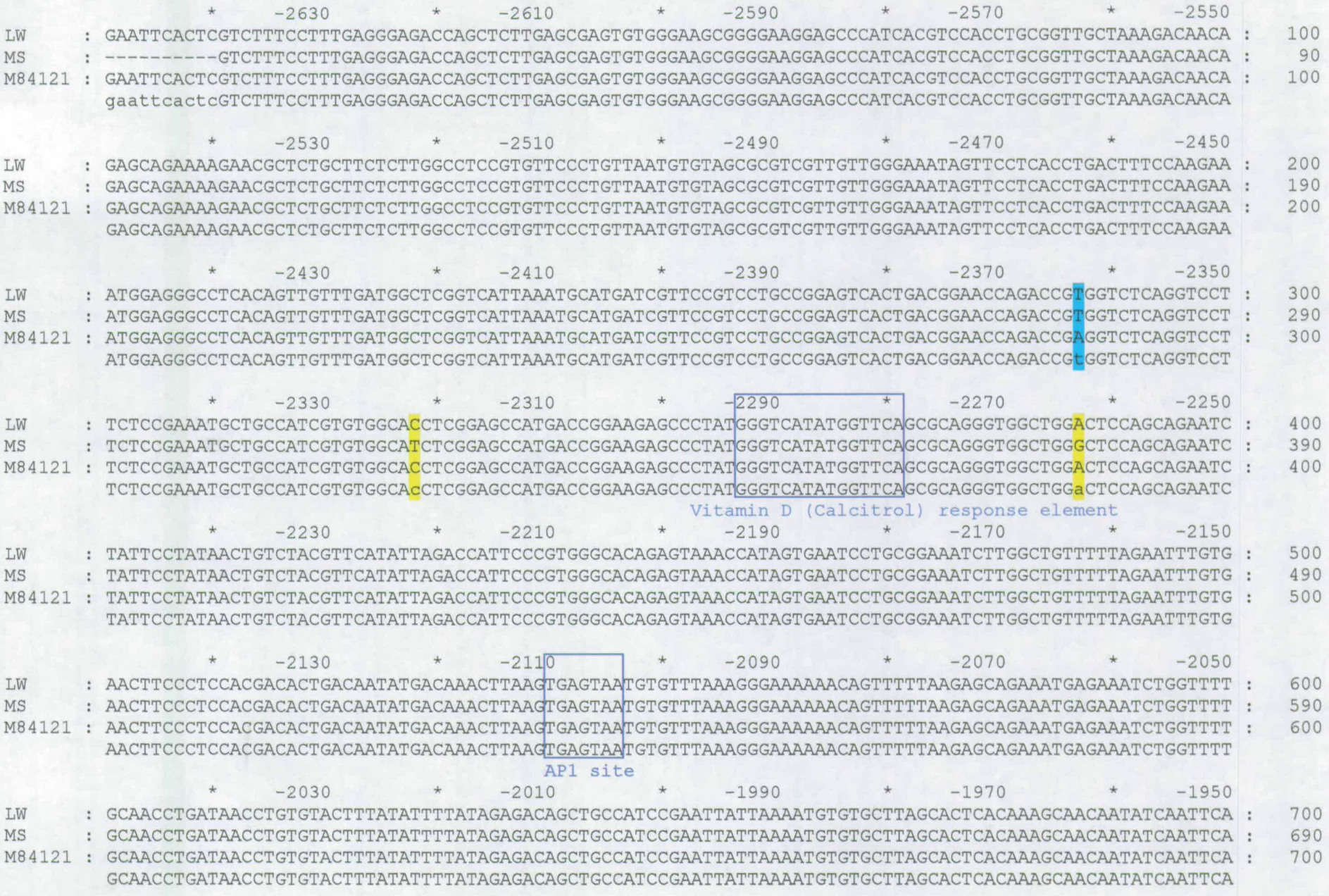
region against the human promoter region (GenBank accession number S78410) also showed large regions of high sequence identity.

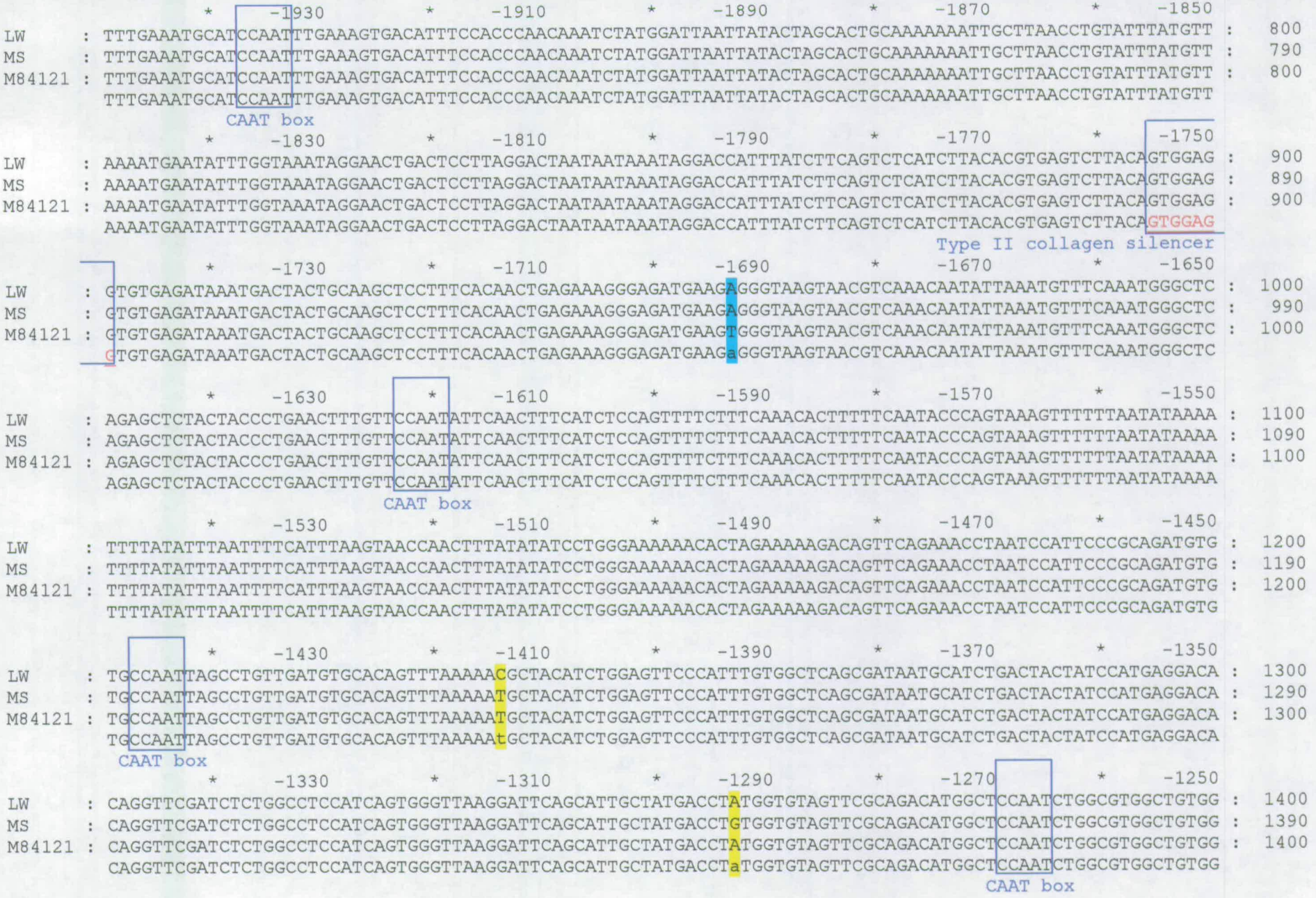
Of the twenty positions where sequence variants were observed between the three sequences (Large White, Meishan and M84121), the published sequence is identical to the Large White sequence at thirteen positions and to the Meishan at only one position. The published sequence differs from both the Large White and Meishan at six positions (highlighted in blue on Figure 5-14). Therefore although the breed origin of the published sequence is not stated, it is most likely to be a Western commercial breed such as the Landrace or Large White. As these variants may be a result of sequence errors they would need to be confirmed from analysis of genomic DNA from several animals of different breeds. These putative sequence variants may indeed represent useful polymorphisms within higher performing Western breeds.

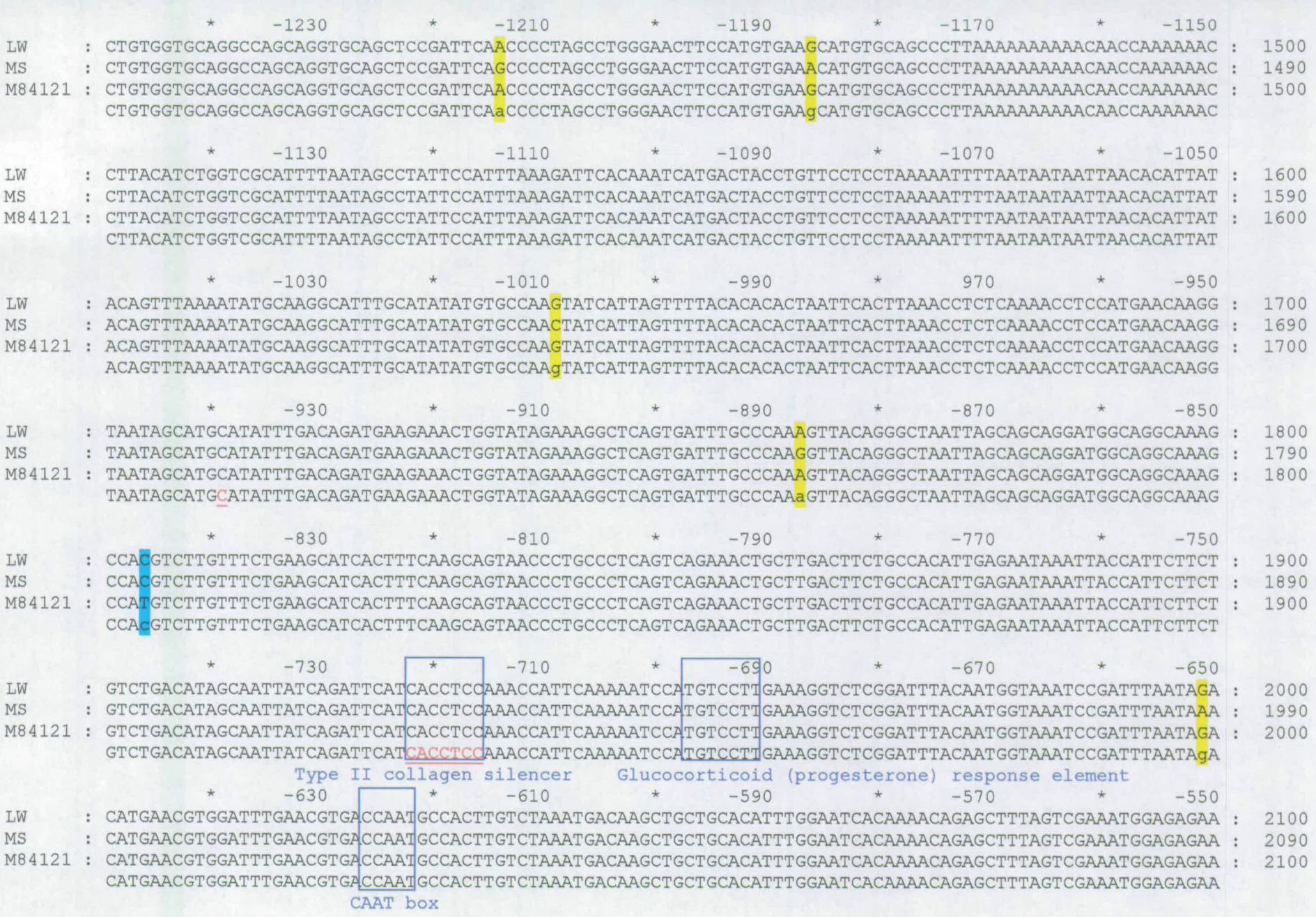
5.3.7.2. Confirmation of SINE repeat element

The nucleotide sequences from each breed origin were aligned with the published porcine sequence of the 3' end of *SPP1* (GenBank accession number AJ237667) (see Figure 5-15). As mentioned in section 5.3.3, Knoll *et al.* (1999) identified variation in the presence of a SINE (PRE-1) sequence within intron 6. It can be clearly seen from the alignment in Figure 5-15 that this SINE element was present in the sequence from the Large White breed, but absent from the Meishan breed sequence. The number of base T repeats in the Large White sequence is longer than the published sequence; this may be as a result of sequencing problems. The polymerase enzyme often has difficulty reading past such sequences and shows slippage on the repeats, which tends to introduce extra stutter bases.

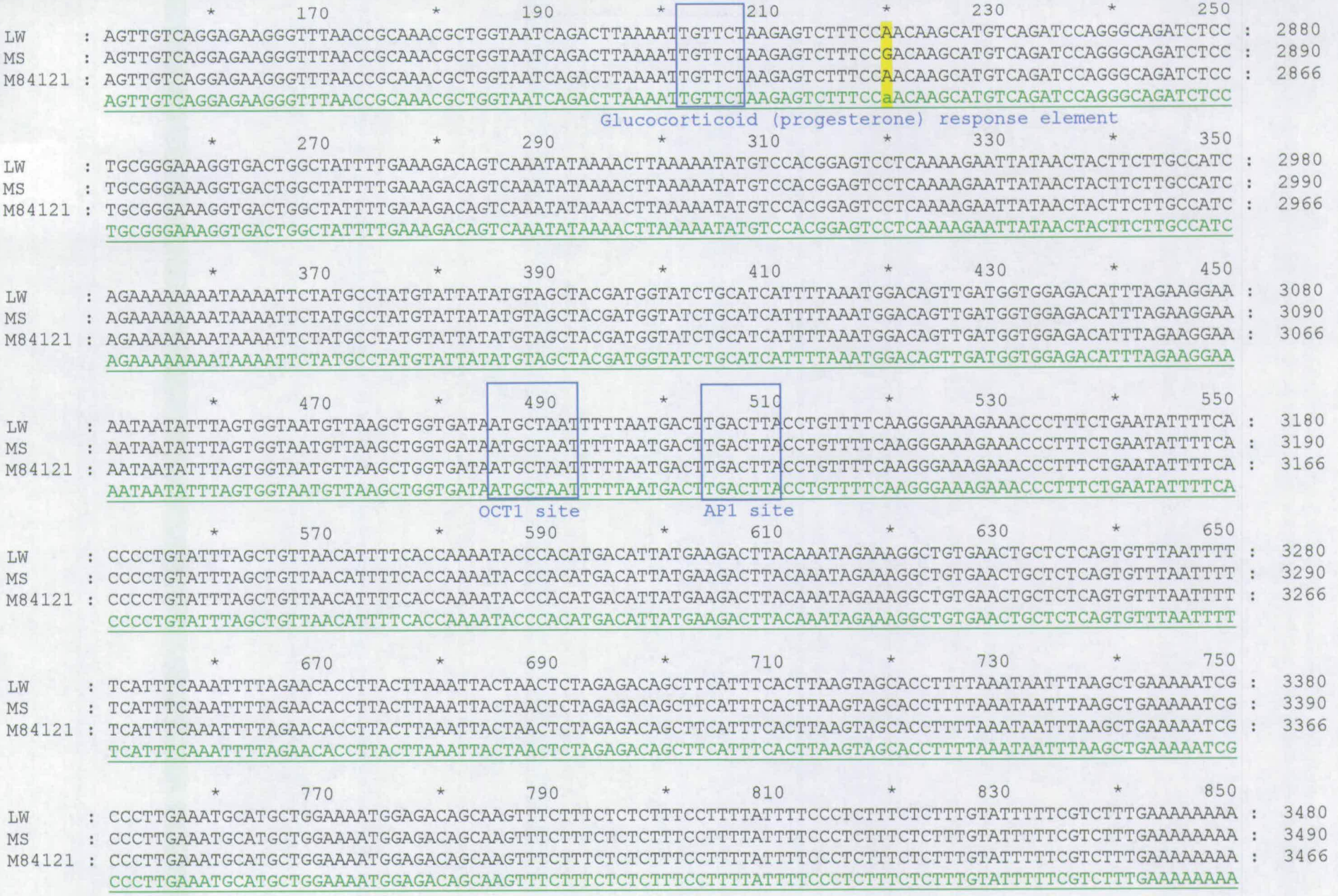
The length of the Large White SINE is 330 bp compared to the published 291 bp sequence and also has the two 14 bp target DNA sequences flanking it (highlighted in pink on Figure 5-15). The published sequence is from a Landrace breed, which is a similar breed to the Large White, and there is in fact only one base throughout the 1295 bp sequence, which appears to be different between these breeds.


Vitamin D (Calcitrol) response element
AP1 site


CAAT box
Type II collagen silencer
CAAT box
CAAT box
CAAT box


* -1230 * -1210 * -1190 * -1170 * -1150
LW : CTGTGGTGCAGGCCAGCAGGTGCAGCTCCGATTCAACCCCTAGCCTGGGAACTTCCATGTGAAGCATGTGCAGCCCTTAAAAAAAAAACAACCAAAAAAC : 1500
MS : CTGTGGTGCAGGCCAGCAGGTGCAGCTCCGATTCAGCCCCTAGCCTGGGAACTTCCATGTGAAACATGTGCAGCCCTTAAAAAAAAAACAACCAAAAAAC : 1490
M84121 : CTGTGGTGCAGGCCAGCAGGTGCAGCTCCGATTCAACCCCTAGCCTGGGAACTTCCATGTGAAGCATGTGCAGCCCTTAAAAAAAAAACAACCAAAAAAC : 1500
CTGTGGTGCAGGCCAGCAGGTGCAGCTCCGATTCAaCCCCTAGCCTGGGAACTTCCATGTGAAgCATGTGCAGCCCTTAAAAAAAAAACAACCAAAAAAC
* -1130 * -1110 * -1090 * -1070 * -1050
LW : CTTACATCTGGTCGCATTTTAATAGCCTATTCCATTTAAAGATTCACAAATCATGACTACCTGTTCCTCCTAAAAATTTTAATAATAATTAACACATTAT : 1600
MS : CTTACATCTGGTCGCATTTTAATAGCCTATTCCATTTAAAGATTCACAAATCATGACTACCTGTTCCTCCTAAAAATTTTAATAATAATTAACACATTAT : 1590
M84121 : CTTACATCTGGTCGCATTTTAATAGCCTATTCCATTTAAAGATTCACAAATCATGACTACCTGTTCCTCCTAAAAATTTTAATAATAATTAACACATTAT : 1600
CTTACATCTGGTCGCATTTTAATAGCCTATTCCATTTAAAGATTCACAAATCATGACTACCTGTTCCTCCTAAAAATTTTAATAATAATTAACACATTAT
* -1030 * -1010 * -990 * 970 * -950
LW : ACAGTTTAAAATATGCAAGGCATTTGCATATATGTGCCAAGTATCATTAGTTTTACACACACTAATTCACTTAAACCTCTCAAAACCTCCATGAACAAGG : 1700
MS : ACAGTTTAAAATATGCAAGGCATTTGCATATATGTGCCAACTATCATTAGTTTTACACACACTAATTCACTTAAACCTCTCAAAACCTCCATGAACAAGG : 1690
M84121 : ACAGTTTAAAATATGCAAGGCATTTGCATATATGTGCCAAGTATCATTAGTTTTACACACACTAATTCACTTAAACCTCTCAAAACCTCCATGAACAAGG : 1700
ACAGTTTAAAATATGCAAGGCATTTGCATATATGTGCCAAgTATCATTAGTTTTACACACACTAATTCACTTAAACCTCTCAAAACCTCCATGAACAAGG
* -930 * -910 * -890 * -870 * -850
LW : TAATAGCATGCATATTTGACAGATGAAGAAACTGGTATAGAAAGGCTCAGTGATTTGCCCAAAGTTACAGGGCTAATTAGCAGCAGGATGGCAGGCAAAG : 1800
MS : TAATAGCATGCATATTTGACAGATGAAGAAACTGGTATAGAAAGGCTCAGTGATTTGCCCAAGGTTACAGGGCTAATTAGCAGCAGGATGGCAGGCAAAG : 1790
M84121 : TAATAGCATGCATATTTGACAGATGAAGAAACTGGTATAGAAAGGCTCAGTGATTTGCCCAAAGTTACAGGGCTAATTAGCAGCAGGATGGCAGGCAAAG : 1800
TAATAGCATGCATATTTGACAGATGAAGAAACTGGTATAGAAAGGCTCAGTGATTTGCCCAAaGTTACAGGGCTAATTAGCAGCAGGATGGCAGGCAAAG
* -830 * -810 * -790 * -770 * -750
LW : CCACGTCTTGTTTCTGAAGCATCACTTTCAAGCAGTAACCCTGCCCTCAGTCAGAAACTGCTTGACTTCTGCCACATTGAGAATAAATTACCATTCTTCT : 1900
MS : CCACGTCTTGTTTCTGAAGCATCACTTTCAAGCAGTAACCCTGCCCTCAGTCAGAAACTGCTTGACTTCTGCCACATTGAGAATAAATTACCATTCTTCT : 1890
M84121 : CCATGTCTTGTTTCTGAAGCATCACTTTCAAGCAGTAACCCTGCCCTCAGTCAGAAACTGCTTGACTTCTGCCACATTGAGAATAAATTACCATTCTTCT : 1900
CCAcGTCTTGTTTCTGAAGCATCACTTTCAAGCAGTAACCCTGCCCTCAGTCAGAAACTGCTTGACTTCTGCCACATTGAGAATAAATTACCATTCTTCT
* -730 * -710 * -690 * -670 * -650
LW : GTCTGACATAGCAATTATCAGATTCATCACCTCCAAACCATTCAAAAATCCATGTCCTTGAAAGGTCTCGGATTTACAATGGTAAATCCGATTTAATAGA : 2000
MS : GTCTGACATAGCAATTATCAGATTCATCACCTCCAAACCATTCAAAAATCCATGTCCTTGAAAGGTCTCGGATTTACAATGGTAAATCCGATTTAATAAA : 1990
M84121 : GTCTGACATAGCAATTATCAGATTCATCACCTCCAAACCATTCAAAAATCCATGTCCTTGAAAGGTCTCGGATTTACAATGGTAAATCCGATTTAATAGA : 2000
GTCTGACATAGCAATTATCAGATTCATCACCTCCAAACCATTCAAAAATCCATGTCCTTGAAAGGTCTCGGATTTACAATGGTAAATCCGATTTAATAgA
Type II collagen silencer
Glucocorticoid (progesterone) response element
* -630 * -610 * -590 * -570 * -550
LW : CATGAACGTGGATTTGAACGTGACCAATGCCACTTGTCTAAATGACAAGCTGCTGCACATTTGGAATCACAAAACAGAGCTTTAGTCGAAATGGAGAGAA : 2100
MS : CATGAACGTGGATTTGAACGTGACCAATGCCACTTGTCTAAATGACAAGCTGCTGCACATTTGGAATCACAAAACAGAGCTTTAGTCGAAATGGAGAGAA : 2090
M84121 : CATGAACGTGGATTTGAACGTGACCAATGCCACTTGTCTAAATGACAAGCTGCTGCACATTTGGAATCACAAAACAGAGCTTTAGTCGAAATGGAGAGAA : 2100
CATGAACGTGGATTTGAACGTGACCAATGCCACTTGTCTAAATGACAAGCTGCTGCACATTTGGAATCACAAAACAGAGCTTTAGTCGAAATGGAGAGAA
CAAT box


cAMP response
GC box, Egr-1 site
CAAT box
OCT-1 site
CAAT box
E4TF1 site
AP1 site
GC box
TATA box


Glucocorticoid (progesterone) response element
OCT1 site
AP1 site

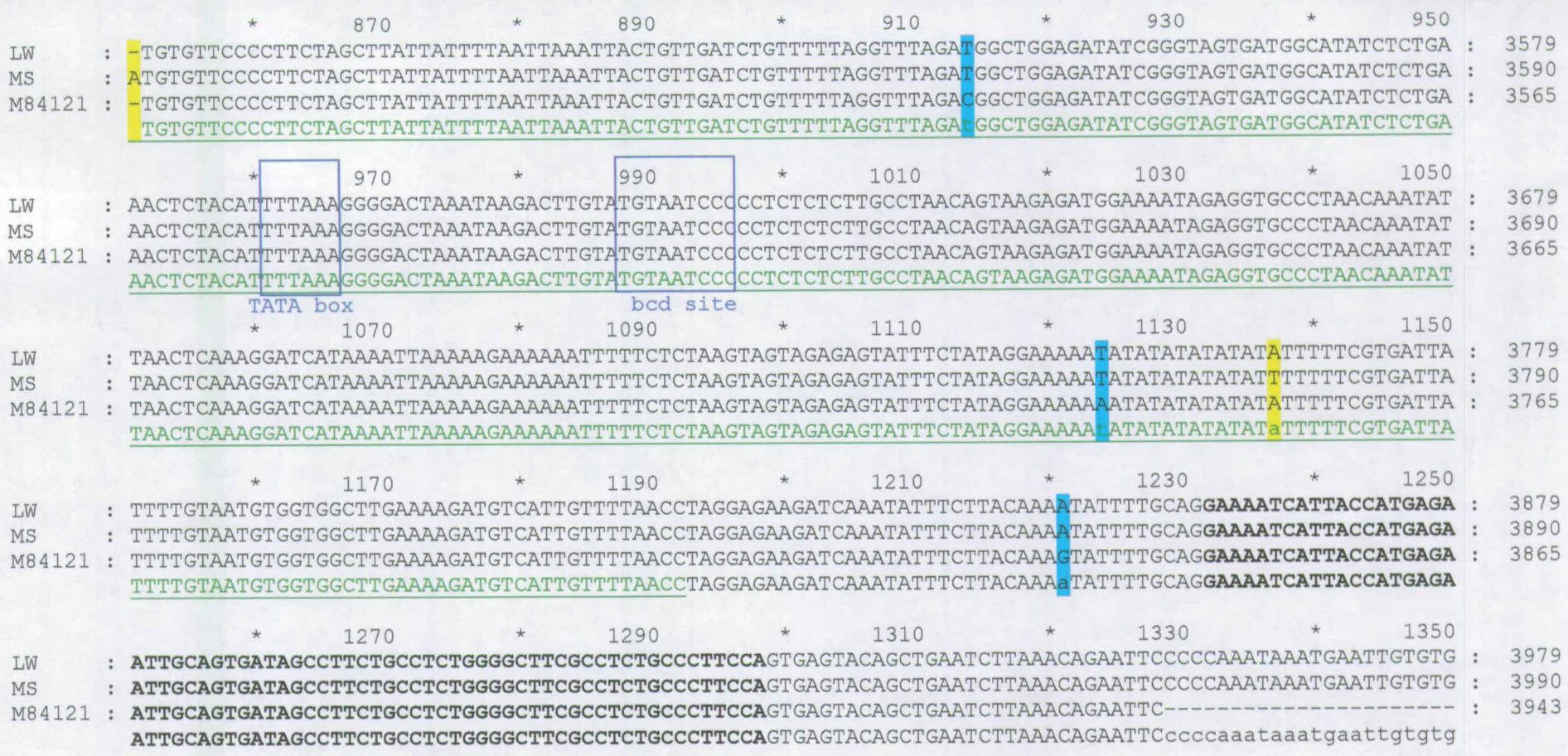
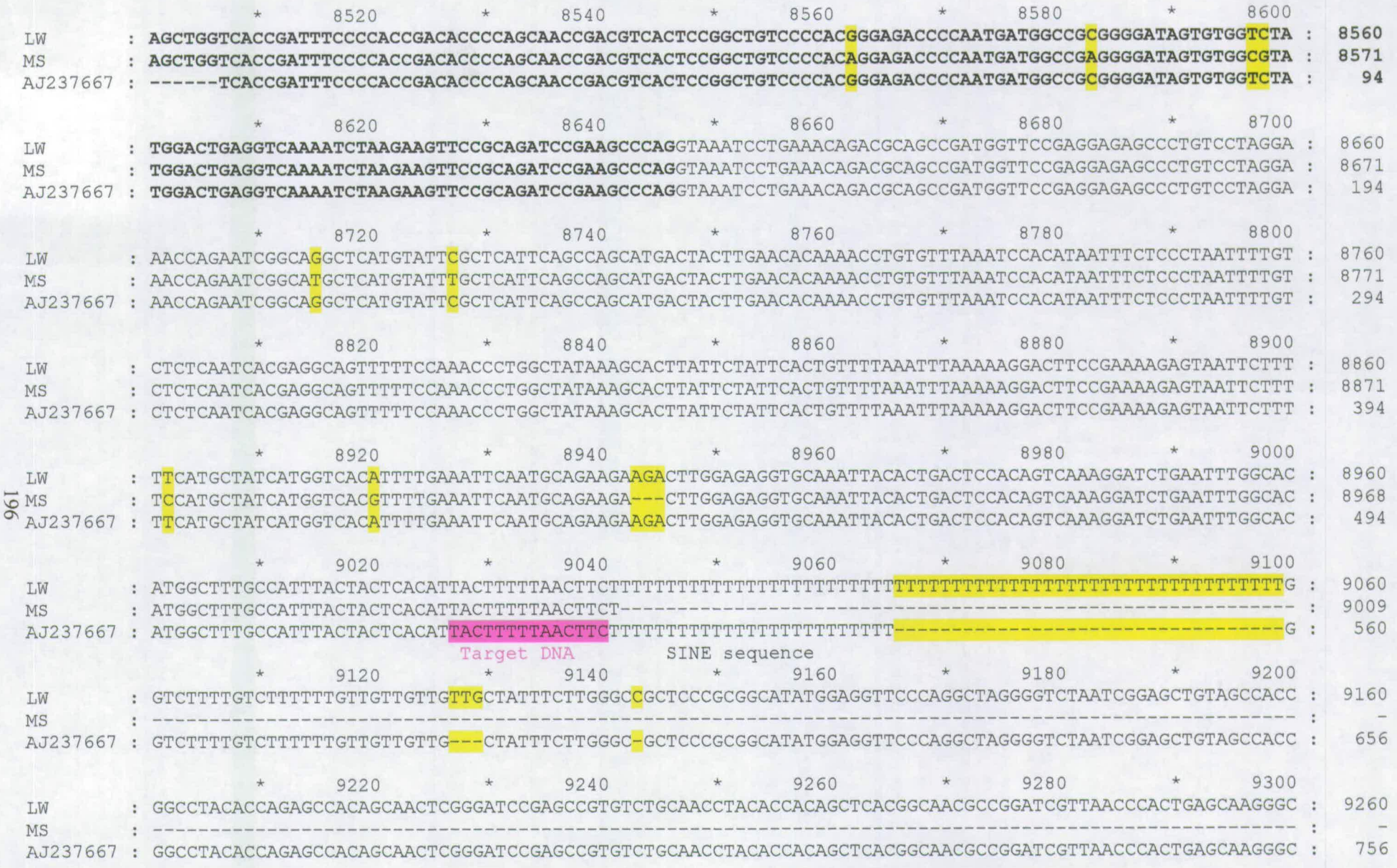


Figure 5-14 Alignment of Meishan (MS) and Large White (LW) sequences with published promoter region of porcine *SPP1* showing consensus regulatory regions (blue boxes) and region of high promoter activity (green text on consensus sequence) and negative regulatory regions (red text on consensus sequence). Sequence variants are highlighted in yellow where the MS sequence varies from the LW and the published sequence and in blue where the published sequence varies from the MS and LW sequences. Exon regions are shown in bold text. Position 0 is the transcription start site.


Target DNA
SINE sequence

```
                    *      9320          *      9340          *      9360          *      9380          *      9400
LW        : AGGGACCGAACCCGCAACCTCATGGTTCCTAGTCGGATTCGTTAACCACTGCGCCACGATGGGAACTCCTACTTTTTAACTTCTGATTCCGTTTCCCTTT :  9360
MS        : ---------------------------------------------------------------------------------GATTCCGTTTCCCTTT :  9025
AJ237667  : AGGGACCGAACCCGCAACCTCATGGTTCCTAGTCGGATTCGTTAACCACTGCGCCACGATGGGAACTCCTACTTTTTAACTTCTGATTCCGTTTCCCTTT :   856
                                                                  Target DNA
                    *      9420          *      9440          *      9460          *      9480          *      9500
LW        : ATGTAAAACAGACGAGACCCGCCTCGTAGATTTCCTTGCATAGATAATGGATGTGCCTATTAAAGTAAAGGAGATGACAGAGTAAAAGGACCTGGTGAAA :  9460
MS        : ATGTAAAACAGACGAGACCCGCCTCGTAGATTTCCTCGCATAGATAATGGATGTGCCTATTAAAGTAAAGGAGATGACAGAGTAAAAGGACCTGGTGAAA :  9125
AJ237667  : ATGTAAAACAGACGAGACCCGCCTCGTAGATTTCCTTGCATAGATAATGGATGTGCCTATTAAAGTAAAGGAGATGACAGAGTAAAAGGACCTGGTGAAA :   956

                    *      9520          *      9540          *      9560          *      9580          *      9600
LW        : AAAAAAAAAGTAACACACTGGTATTATTATTATTACTTCAGCCACATTTAGACATTTCTGTAGAATACACTGGCCTGAAAGTTAGAAAGCAGAAAAGAGC :  9560
MS        : AAAAAAAAAGTAACACACTGGTATTATTATTATTACTTCAGCCACATTTAGACATTTCTGTAGAATACACTGGCCTGAAAGTTAGAAAGCAGAAAAGAGC :  9225
AJ237667  : AAAAAAAAAGTAACACACTGGTATTATTATTATTACTTCAGCCACATTTAGACATTTCTGTAGAATACACTGGCCTGAAAGTTAGAAAGCAGAAAAGAGC :  1056

                    *      9620          *      9640          *      9660          *      9680          *      9700
LW        : TACAGTGTTCCCATCCCTGGCTGTTCATTTAATTCTTCTCCCATTTTTGCTGTGATTCAGCAGCTGGATGCCACAGAGGAAGACCTCACGTCACATGTGG :  9660
MS        : TACAGTGTTCCCATCCCTGGCTGTTCATTTAATTCTTCTCCCATTTTTGCTGTGATTCAGCAGCTGGATGCCACAGAGGAAGACCTCACGTCACATGTGG :  9325
AJ237667  : TACAGTGTTCCCATCCCTGGCTGTTCATTTAATTCTTCTCCCATTTTTGCTGTGATTCAGCAGCTGGATGCCACAGAGGAAGACCGCACGTCACATGTGG :  1156

                    *      9720          *      9740          *      9760          *      9780          *      9800
LW        : AAAGTGAGGAGACGGATGGTACCCCCAAGGCCATCCTCGTTGCCCAGCGCCTGCACGTGGCTTCTGACTTGGACAGCCAAGAGAAGGACAGTCAGGAGAC :  9760
MS        : AAAGTGAGGAGACGGATGGTACCCCCAAGGCCATCCTTGTTGCCCAGCGCCTGCACGTGGCTTCTGACTTGGACAGCCAAGAGAAGGACAGTCAGGAGAC :  9425
AJ237667  : AAAGTGAGGAGACGGATGGTACCCCCAAGGCCATCCTCGTTGCCCAGCGCCTGCACGTGGCTTCTGACTTGGACAGCCAAGAGAAGGACAGTCAGGAGAC :  1256

                    *      9820          *      9840          *      9860          *      9880          *      9900
LW        : GAGTCAGCCGGATGACCGCAGTGTGGAAACCCGCAGCCAGGAGCAGTCCAAAGAATACACGATCAAGACCTATGATGGGAGCAATGAGCATTCCAATGTG :  9860
MS        : GAGTCAGCCGGATGACCGCAGTGTGGAAACCCGCAGCCAGGAGCAGTCCAAAGAATACACGATCAAGACCTATGATGGGAGCAATGAGCATTCCAATGTG :  9525
AJ237667  : GAGTCAGCCGGATGACCGCAGTGTGGAAACCCGCAGCCA---------------------------------------------------------- :  1295
```

Figure 5-15 Alignment of Meishan (MS) and Large White (LW) sequences with published 3' end of porcine *SPP1*. Sequence variants are highlighted in yellow where the MS sequence varies from the LW and the published sequence and in blue where the published sequence varies from the MS and LW sequences. The two 14 bp target DNA sequences flanking either side of the SINE are highlighted in pink.

5.3.8. Homology searches

5.3.8.1. Nucleotide sequence

The GenBank, EMBL, DDBJ and PDB databases were searched with the complete nucleotide sequence for the Large White copy of the gene using BLAST. The only match for the entire length of the gene sequence was with human sequence (GenBank accession number D14813). For other species, at present only cDNA sequences rather than full gene sequences are available in the public databases. The human gene structure appears to be a similar size and structure to that of the pig. To calculate sequence identity between sequences, pairwise alignments were made using the GAP programme within GCG, the published sequences were imported into GCG using the xfetch option. The Large White and Meishan sequences were found to be 99.123 % identical. The Large White sequence was 44.468 % identical to the human and the Meishan sequence had 44.189 % similarity.

5.3.8.2. Amino acid sequence

As with the nucleotide sequence, the coding sequence of the Large White was BLAST searched against GenBank CDS translations, PDB, SwissProt, PIR and PRF databases. There were several protein sequences from mammalian species found to have a high sequence identity (the SwissProt accession numbers are shown in brackets after the species names). The sequences for *Sus scrofa* (P14287), *Homo sapiens* (P10451), *Mus musculus* (P10923), *Rattus norvegicus* (P08721), *Oryctolagus cuniculus* (P31097), *Ovis aries* (Q9XSY9) and *Bos taurus* (P31096) were then aligned with the Large White and Meishan sequences, in order to identify regions of high sequence identity between species (see Figure 5-16). The regions that are conserved across species are very likely to be functionally important to the protein, for example the RGD binding site.

In addition pairwise alignments were carried out, using the GAP programme within GCG, to discover the level of identity between the amino acid sequences from different species (see Figure 5-17).


EXON 2
EXON 3
EXON 4
EXON 5
EXON 6
LW : MRIAVIAFCLWGFASALPVKQTNSGSSEEKLLSNKYTDAVATLLKPDPSQKQTFLAPQNTISSEETDDFKQETLPSKSNESPEQT : 85
Pig : MRIAVIAFCLWGFASALPVKQTNSGSSEEKLLSNKYTDAVATLLKPDPSQKQTFLAPQNTISSEETDDFKQETLPSKSNESPEQT : 85
MS : MRIAVIAFCLWGFASALPVKQTNSGSSEEKLLSNKYTDAVATLLKPDPSQKQTFLAPQNTISSEETDDFKQETLPSKSNESPEQT : 85
Human : MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWLNPDPSQKQNLLAPQNAVSSEETNDFKQETLPSKSNESHDHM : 85
Rabbit : MRIAVICFCLLGMAYALPVKHADSGSSEEKQLYHKHPDALATWLNPDPSQKQNLLTPQNAMSSEEKDDLKQETLPSKSIESHDHM : 85
Mouse : MRLAVICFCLFGIASSLPVKVTDSGSSEEKL-YSLHPDPIATWLVPDPSQKQNLLAPQNAVSSEEKDDFKQETLPSNSNESHDHM : 84
Rat : MRLAVVCFCLFGLASCLPVKVAEFGSSEEKAHYSKHSDAVATWLKPDPSQKQNLLAPQNSVSSEETDDFKQETLPSNSNESHDHM : 85
Cattle : MRIAVICFCLLGIASALPVKPTSSGSSEEKQLNNKYPDAVAIWLKPDPSQKQTFLTPQNSVSSEETDDNKQNTLPSKSNESPEQT : 85
Sheep : MRIAVICFCLLGIASALPVKPTSSGSSEEKQLNNKYPDAVATWLKPDPSQKQTFLEPQNSVSSEETDDNKQNTLPSKSNESPEQT : 85
signal peptide
start secreted protein
Human: Exon 5 missing in
alternative splice variant
Potential Asn-linked glycosylation site
LW : DDVDDDDDEDHVDSRDT----DSEEADHADDADRSDESHHSDESDELVTDFPTDTPATDV-TPAVPTGDPNDGRGDSVVYGL-RS : 164
Pig : DDVDDDDDEDHVDSRDT----DSEEADHADDADRSDESHHSDESDELVTDFPTDTPATDV-TPAVPTGDPNDGRGDSVVYGL-RS : 164
MS : DDVDDDDDEDHVDSRDT----DSEEADHTDDADRSDESHHSDESDELVTDFPTDTPATDV-TPAVPTGDPNDGRGDSVAYGL-RS : 164
Human : DDMDDEDDDDHVDSQDSIDSNDSDDVDDTDDSHQSDESHHSDESDELVTDFPTDLPATEVFTPVVPTVDTYDGRGDSVVYGL-RS : 169
Rabbit : DDIDEDEDDDHVDNRDS---NESDDADHPDDSHHSDESHQSDESDE-VTVYPTEDAATTVFTEVVPTVETYDGRGDSVAYRLKRS : 166
Mouse : DDDDDDDDDDG----------DHAESEDSVDSDESDESHHSDESDETVTA----STQADTFTPIVPTVDVPNGRGDSLAYGL-RS : 154
Rat : DDDDDDDDD-G----------DHAESEDSVNSDESDESHHSDESDESFTA----STQADVLTPIAPTVDVPDGRGDSLAYGL-RS : 154
Cattle : DDLDDDDDNSQDVNSN-----DSDDAETTDDPDHSDESHHSDESDE--VDFPTDIPTIAVFTPFIPTESANDGRGDSVAYGL-KS : 162
Sheep : DDLDDDDENSQEVNSD-----DSDDAETPDDSDHSNESHHSDESDE--ADFPTDIPTIAVFTPPFPTESTNDGRGDSVAYGL-KS : 162
Poly aspartate region(mediate hydroxyapatite binding)
RGD binding site

Figure 5-16 Alignment of amino acid sequence of *SPP1* from several mammalian species with Meishan (MS) and Large White (LW) sequences. Functional regions are highlighted with a blue box. The exons encoding these sequences are noted.

Key:

Primary align (sequences are conserved across 100% of species)

Secondary align (sequences are conserved across 80% of species)

Tertiary align (sequences are conserved across 60% of species)

▭ = amino acid differences between Meishan and Large White sequence

Figure 5-17 Pairwise amino acid sequence similarity of Meishan and Large White sequences with other mammalian species (shown as a percentage)

	Meishan	Large White	Pig	Human	Cow	Sheep	Rabbit	Rat	Mouse
Meishan									
Large White	99.01								
Pig (P14287)	99.01	100.00							
Human (P10451)	71.62	71.29	71.29						
Cattle (P31096)	72.92	71.84	71.84	69.78					
Sheep (Q9XSY9)	70.76	70.04	70.04	67.99	91.73				
Rabbit (P31097)	65.56	64.90	64.90	69.68	61.15	60.43			
Rat (P08721)	63.70	62.67	62.67	68.79	61.11	61.11	61.41		
Mouse (P10923)	61.81	61.11	61.11	68.84	62.74	60.08	60.88	85.32	

5.3.9. Secondary structure prediction

The PIX analysis from the HGMP resource centre (http://www.hgmp.mrc.ac.uk) was used to investigate the structure of the protein sequence from the Meishan and from the Large White. Secondary structure prediction was carried out using the DSC (Discrimination of protein Secondary structure Class) program (King and Sternberg, 1996), which is claimed to be around 70 % accurate. The output is shown in Figure 5-18.

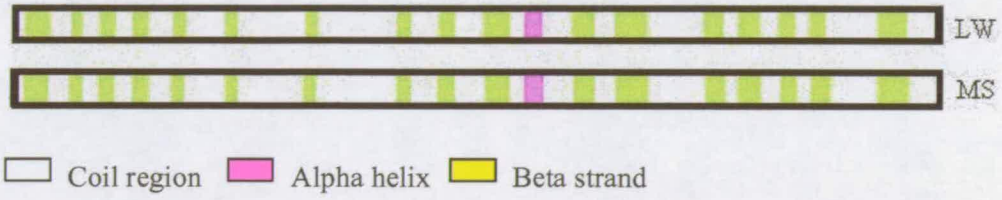


Figure 5-18 Secondary structure prediction of Large White (LW) and Meishan (MS) SPP1 protein sequences

The secondary folding was predicted to be identical for the sequences from both breeds. The valine to alanine amino acid change encoded in exon 6 lies within a beta strand, however the change does not alter the predicted secondary structure. The other two amino acid changes lie within the coil region. The PIX analysis also classifies the amino acids within the sequence. Table 5-3 shows the nature of the amino acid side chains for the alternative amino acids present for the three variants.

Table 5-3 Amino acid properties of variants between Large White (LW) and Meishan (MS) sequences

	Residue 110 (exon 6)		Residue 159 (exon 6)		Residue 262 (exon 7)	
	Alanine (LW)	Threonine (MS)	Valine (LW)	Alanine (MS)	Proline (LW)	Serine (MS)
Acidity	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
Polarity	Neutral	Polar	Neutral	Neutral	Neutral	Polar
Hydrophobicity	Hydro-phobic	Hydro-phobic	Hydro-phobic	Hydro-phobic	Hydro-phobic	Hydro-phobic
Aromaticity	Neutral	Neutral	Aliphatic	Neutral	Neutral	Neutral
Position	Buried	Surface	Buried	Buried	Surface	Surface
Charge	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
Size	Tiny	Small	Small	Tiny	Small	Tiny

5.4. Discussion

In the QTL study (chapter 2) litter size and prenatal survival rates were recorded for F_2 generation cross-bred sows and an association between the telomeric region of chromosome 8 and a difference in the performance of animals with alleles of Meishan and Large White breed origin for these traits was found. As previously mentioned, *SPP1* maps within the 95 % confidence interval of this QTL.

SPP1 is an acidic phosphorylated glycoprotein involved in controlling trophoblast elongation and implantation during the peri-implantation period and is a strong candidate gene for control of the differences in prenatal survival levels observed between the Meishan and Large White breeds. The Meishan uterine environment restricts the development and elongation rate of the conceptus during the peri-implantation period (days 12-18 of gestation) and consequently allows an increased number to survive this critical period of loss (Biensen *et al.*, 1999 and (Wilson et al., 1998). See section 1.4.2 for more detail on the differences in the control of prenatal survival between the Meishan and Large White breeds.

The native 70 kDa form of SPP1 is degraded to the biologically active 45 kDa form, which is then released as part of the histotroph secretion from the endometrium. This active form of SPP1 has a high affinity for $\alpha v\beta 3$ integrin (vitronectin) receptors expressed by the trophectoderm (the outer layer of the blastocyst) and the uterus during the implantation window (Johnson et al., 2000).

The binding of SPP1 to the vitronectin receptors on the trophectoderm and uterus stimulates changes in morphology of the trophectoderm and extra-embryonic endoderm that result in the cytoskeletal reorganisation and elongation of the conceptus. It also induces adhesion and cell signalling between the luminal epithelium and trophectoderm essential for attachment, superficial implantation and placentation (Johnson et al., 1999b; Johnson et al., 2000).

In pigs, an increased expression of *SPP1* mRNA has been shown in the uterine luminal epithelium, in regions of close proximity of conceptus tissue, after day 15 of

gestation (peri-implantation period) and has been shown to result in integrin activation and the accumulation of the cytoskeletal molecules required to form the "focal adhesions" for adhesion and signalling between the conceptus and the uterus (Garlow et al., 2002). Interestingly porcine conceptuses were not shown to express *SPP1* mRNA, however the protein was found at the conceptus trophectoderm as well as the uterine luminal epithelium (Garlow et al., 2002).

The release of progesterone and interferon-tau (the maternal recognition of pregnancy signal) from the trophoblast, during the peri-implantation period, have been shown to result in this increased expression of *SPP1* mRNA from the luminal epithelium of the uterine endometrium in humans (Omigbodun et al., 1997), mice (Nomura et al., 1988) and pigs (Garlow et al., 2002), from the glandular epithelium of the uterus and decidualising stroma of baboons (Fazleabas et al., 1997) and from the glandular epithelium of ewes (Johnson *et al.*, 2000 and (Johnson et al., 1999b). Indeed the mouse and human *SPP1* genes possess a putative progesterone regulatory element in the 5' flanking region (Craig and Denhardt, 1991; Hijiya et al., 1994), this motif was also seen in the pig sequence (Zhang et al., 1992a). It is also believed that there is an oestrodiol response element in the murine promoter region and Craig and Denhardt (1991) demonstrated that oestradiol as well as progesterone can induce expression of *SPP1* mRNA. This oestradiol response element was not identified in the study to characterise the porcine promoter region of *SPP1* (Zhang et al., 1992a).

From days 12-15 of gestation the Meishan embryos have a smaller number of trophectoderm cells than Yorkshire embryos and as a consequence they secrete less oestradiol to the uterus (Ford, 1997). In response to this lower level of oestradiol from the Meishan embryos, it is possible that the level of *SPP1* expression is reduced in the uterus of the Meishan sows compared to the Large White. The postulated reduced level of SPP1 could result in less trophectoderm elongation in this breed. However it is not entirely clear whether this suggestive reduction in *SPP1* expression would affect implantation. Thus, the importance of this gene in the control of overall variation in prenatal survival levels between the two breeds cannot be clearly ascertained from the literature available. There will of course be other proteins

present in the histotroph secretion from the endometrium, which also control embryo development.

By sequencing a copy of *SPP1* from both breeds it was hoped that variation would be identified, in particular in regulatory regions, that may indicate a change in the control of the level of gene expression or a variation in protein structure and subsequent function between breeds. A relatively high level of variation was identified between the two gene sequences, with an average of one sequence variant per 104 base pairs. As a comparison, Fahrenkrug *et al.* (2002) observed single nucleotide polymorphisms (SNPs) on average every 184 bp, in regions of expressed genomic DNA from crossbred pigs of Western and Chinese origin. SNPs do indeed represent the most abundant form of genetic variation and occur on average at every 100-300 bases in the human genome. Many SNPs have no effect on cell function, but it is believed that some can predispose individuals to disease and influences how different people respond to drugs (SNP fact sheet, http://www.ornl.gov/hgmis/faq/snps.html).

The most obvious difference between the sequences from the Meishan and Large White was where the SNP variant resulted in a change in amino acid sequence and could therefore subsequently result in a change in the structure and properties of the protein. There were found to be three amino acid variants between the predicted SPP1 sequence. The valine to alanine encoded in exon 6 is a relatively conserved change in that the side chains of both amino acids are non-polar and these residues are of the same family and will therefore have very similar properties. In contrast, the alanine to threonine and the proline to serine changes encoded in exons 6 and 7 both represent changes from a non-polar side chain residue, which tends to cluster on the inside of the protein, to an uncharged polar side chain, which is relatively hydrophilic and tends to remain on the outside of the protein (Alberts et al., 2002). Figure 5-16 shows the level of similarity of the SPP1 sequence with other species and it can be seen that of the three amino acid changes the serine residue (position 286 on Figure 5-16) is the most conserved across species. In fact only the Large White and published pig sequence had a proline residue rather than a serine at this

position. Not unexpectedly due the similarity in the structure of each of the three pairs of amino acids, the secondary folding of the protein was predicted to be the same for the two breeds of pig.

The expression of *SPP1* is controlled by complex regulatory systems, which differ between cell types, in some tissues the gene is expressed constitutively and in others expression is switched on and off (Yamamoto et al., 1995). Control of transcription is developmentally regulated in different tissues. Many different growth and differentiation factors as well as hormones and tumour promoters influence expression levels. Also alternative promoters are used in different tissues determining the size of the resulting transcripts (Zhang et al., 1992a).

Craig and Denhardt (1991) suggested that transcription from the mouse *SPP1* promoter is normally repressed or controlled at the posttranscriptional level and that different promoter regions are active in varying tissues. For example the promoter 5' of exon 1 is active in epidermal, fibroblast and osteoblast-like cells and repressed in T cells where the promoter in intron 1 is active (Craig and Denhardt, 1991).

As mentioned previously, it is likely that it is the level of expression of the *SPP1* gene that is important in controlling the rate of trophoblast elongation. Therefore sequence variation between breeds was investigated within the known regulatory regions of the gene. The promoter region of human, mouse and porcine *SPP1* had been characterised by (Craig and Denhardt, 1991; Hijiya et al., 1994; Yamamoto et al., 1995; Zhang et al., 1992a). Figure 5-14 shows that there was no variation between the gene sequences of the two breeds in any of the putative consensus sequences for promoter regulatory regions, including putative progesterone response elements, from the published porcine sequence (Zhang et al., 1992a). However, the main difference in the intron 1 promoter region is the length of the d(TG)n microsatellite repeats, although no evidence was found for a role of this DNA repeat in gene expression levels. Intron 1 contains regions of good promoter activity and coincidently a high level of conservation was seen between the sequences of the two breeds compared to the other introns in the gene sequence (Table 5-3). This could

reflect an evolutionary pressure to retain important regulatory regions and therefore any SNP found within this region could have a significant role in gene expression levels.

The TATA box at position –30 of the transcription start site (Figure 5-14) is believed to function as a positioning factor initiating transcription by RNA polymerase II. TATA boxes are found in the promoter regions of many genes, although the sequence is usually TATAAA and it is not known why the porcine and mouse sequence was found to be TTTAAA by myself and also by Yamamoto *et al.* (1995) and Craig and Denhardt (1991). There is also a VDRE-like sequence at position -2293. It has been reported that there is a short term regulation in osteoblast-like cells by 1,25-$(OH)_2$-vitamin D3 of posttranslational modification of SPP1 which may modify the functional properties of the protein (Safran et al., 1998). Both of these regulatory motifs are conserved across the two breeds.

Transcription from the *SPP1* promoter is normally repressed or controlled at the posttranscriptional level (Yamamoto et al., 1995). SPP1 does indeed have a significant number of potential serine and threonine sites that can be phosphorylated (Safran et al., 1998). Interestingly phosphorylated SPP1 shows cell surface associations and the non-phosphorylated form is not found at the cell surface (Singh et al., 1990). There is also evidence to suggest that a decreased level of phosphorylation leads to reduced binding of SPP1 to those cells containing the $\alpha_v\beta_3$ integrin receptors (Safran et al., 1998). Also SPP1 is rich is aspartic acid residues which can be heavily glycosylated (Hu et al., 1995). I decided that it would be worthwhile investigating the potential posttranslational modification sites within the gene and see whether the sequence of these sites differed between the Meishan and Large White. If posttranslational modification of the protein varies between the breeds it may explain a difference in the level of binding of SPP1 to vitronectin receptors and the subsequent reduction in the rate of trophoblast development observed in the Meishan breed.

In those tissues where SPP1 modification has been shown to occur at a transcriptional level, it appears to be activated by protein kinase C, cAMP and G-proteins (Craig and Denhardt, 1991). Indeed Ashkar *et al.* (1995) demonstrated that a mouse osteosarcoma cell line attaches and spreads on surfaces coated with either golgi rich membrane preparations or with casein kinase II, but not with unphosphorylated or dephosphorylated protein or protein phosphorylation with cAMP or cGMP dependant protein kinase. It therefore appears that it is crucial for SPP1 to be phosphorylated for example with casein kinase II in order to mediate cell matrix interactions, which allow the cell to respond to various stimuli for example cell signalling between the luminal epithelium and the trophectoderm.

Mouse SPP1 contains consensus sequences for various protein kinases, including several sites for casein kinase II and cAMP dependant protein kinases (Ashkar et al., 1993). At least 58 consensus phosphorylation sites for different types of kinases appear to be organised into eight clusters in mouse SPP1 (Ashkar et al., 1993). Not all potential sites will be phosphorylated in any tissue at a given time. The form of the phosphorylated protein depends on the cell type and/or the metabolic state of the cell (Ashkar et al., 1993).

SPP1 from rat bone has been shown to contain 12 phosphoserines and 1 phosphothreonine, with 1 N-linked and 5 or 6 O-linked glycosylations (Prince et al., 1987). Bovine SPP1 contains a total of 41 serines, 17 threonines and 2 tyrosines, giving a total of 60 potential phosphorylation sites (Kerr et al., 1991). Sorensen *et al.* (1995) showed that bovine SPP1 from milk contains a total of 28 phosphorylation sites and 3 O-glycosylation sites (these sites are marked on Figure 5-19). Sorensen and Petersen (1994) stated that Ser–X–Glu/Ser motifs tend to be phosphorylated by mammary gland casein kinase (MGCK) and Ser–X-X-Glu/Ser is the recognition sequence for casein kinase II (CKII).

It can be seen from Figure 5-16 that the amino acid change from a proline to serine (residue 286 on Figure 5-16) is in fact within a recognition sequence for casein kinase II. The Meishan sequence is Ser-X-X-Ser and the first serine will be

phosphorylated, whereas the Large White sequence is Pro-X-X-Ser and therefore will not be phosphorylated. In bovine milk SPP1, two serines in a MGCK recognition sequence and two serines in a CKII recognition motif (including the amino acid change at residue 286) were found not to be phosphorylated, although Sorensen *et al.* (1995) suggest that they may well be phosphorylated in different tissues and at various stages of development.

Alignment of the bovine amino acid sequence with other species (Figure 5-16 and Figure 5-19) shows that the milk phosphorylation sites are mostly conserved, 16 out of the 28 phosphorylation sites are conserved across all species (pigs of Large White and Meishan breed, human, rabbit, mouse, rat, cattle and sheep). This high degree of conservation of serines and threonines suggests that the phosphorylation of SPP1 at specific sites is essential for the function of the protein (Sorensen et al., 1995). In addition the SNP site within exon 7 encoding either a proline or serine is also highly conserved across species, in that a serine is present in the protein sequence of all species except the Large White and Landrace breeds of pig (Figure 5-16). All these conserved sites across species include a strong evolutionary pressure to conserve these sites.

The sequence comparisons of SPP1 from Meishan and Large White breed origin revealed no apparent differences in the secondary folding of the protein or in the known regulatory regions of the gene within the direct promoter regions. However as mentioned, post-transcriptional modification plays an important role in the level of *SPP1* expression. The gene does indeed have a significant number of conserved potential serine and threonine sites, which can be phosphorylated. This high degree of conservation of serines and threonines suggests that the phosphorylation of SPP1 at specific sites is essential for the function of the protein within various tissues and stages of development. As mentioned, the amino acid change from a proline to serine is within a recognition sequence for casein kinase II, such that the Meishan sequence would be phosphorylated and the Large White sequence would not be. It would be interesting to discover whether this site is indeed phosphorylated by CKII in the endometrial tissues of the pig.

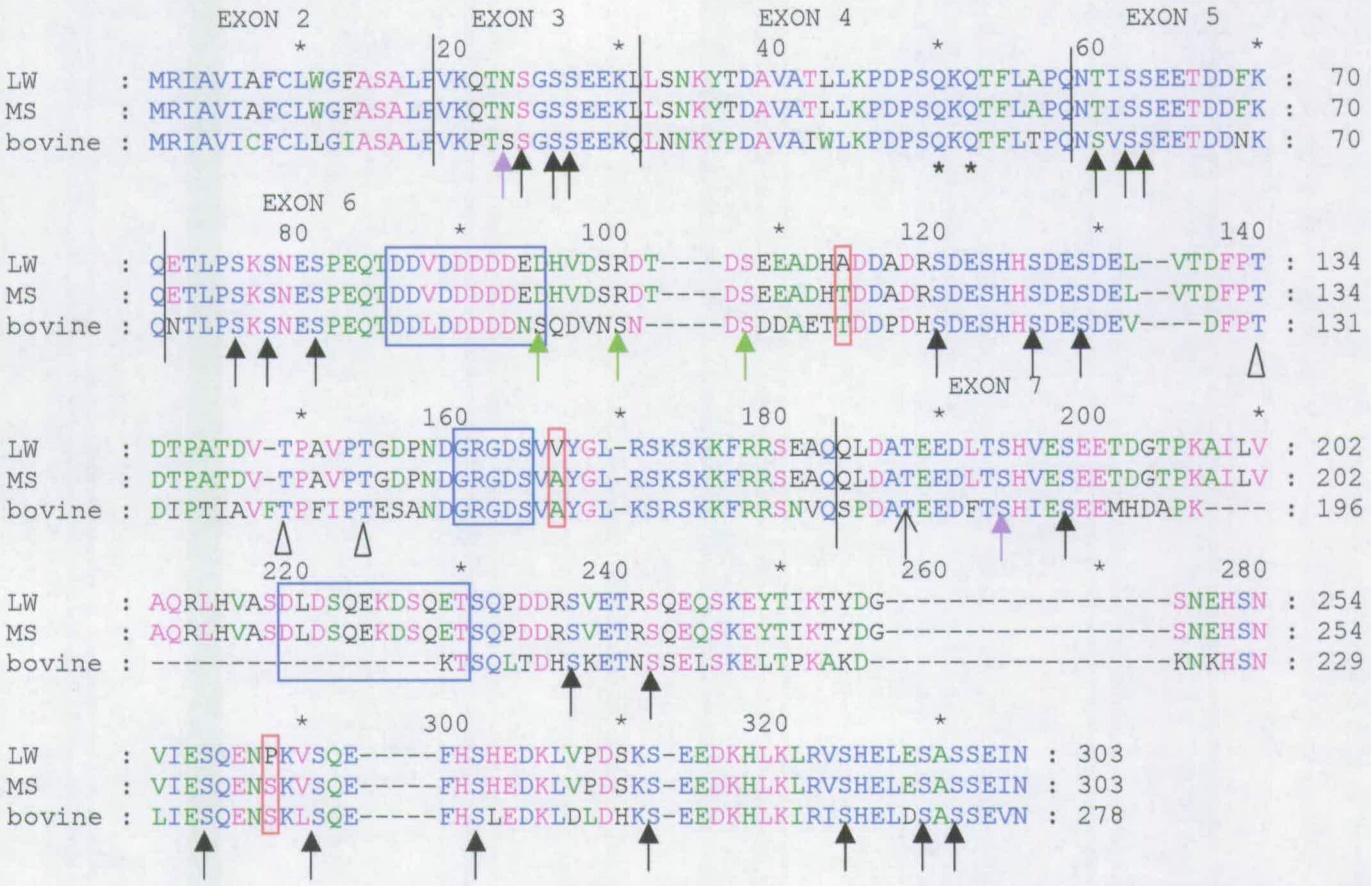


Figure 5-19 Amino acid alignment taken from figure 5-16. Sites of phosphorylation and glycosylation in bovine milk SPP1 are marked below the sequence. The exons encoding the sequences are noted.

KEY: = Phosphorylation of serine within Ser-X-X-Glu/Ser motif (n=2), = Phosphorylation of serine within Ser-X-Asp motif (n=3)

= Phosphorylation of serine within Ser-X-Glu/Ser motif (n=22), = *O* -glycosylation of threonine (n=3)

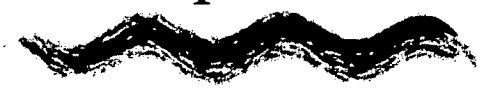
= Phosphorylation of threonine within Thr-X-Glu/Ser motif (n=1) * = tranglutaminase-reactive glutamines (n=2)

SPP1 is expressed on the surface of the luminal epithelial cell of the endometrium and unlike unphosphorylated protein, phosphorylated SPP1 has indeed been shown to have cell surface associations. There is also evidence to suggest that a decreased level of phosphorylation leads to reduced binding of SPP1 to those cells containing the $\alpha v\beta 3$ integrin receptors. However this evidence contradicts the finding that the Meishan SPP1 protein will be phosphorylated at the CKII site proposed in exon 7, whereas the Large White protein would not be. If the decreased level of phosphorylation had been observed in the Meishan and not the Large White, this would correlate with reduced binding of SPP1 to vitronectin receptors and the subsequent decreased elongation of the Meishan conceptuses. However the relationship between phosphorylation of SPP1, receptor activation and reduced elongation of the conceptus may well be more complex, it is only really possible to state that there may be a difference in exchange of information between the implantation site and the conceptus, and some of that could as easily be growth retarding as well as growth enhancing.

The literature suggests that it is the mother's genotype and not the conceptus that controls the significant difference in the level of peri-implantation loss between breeds (see section 1.4.2). Indeed an association was found between the genotypes of F_2 sows and litter size and prenatal survival levels for these animals in the QTL study of chapter 2. In addition SPP1 mRNA is only expressed from the uterine endometrium and not by the conceptus (Garlow et al., 2002).

The main putative sequence variants of interest, for example the SINE and the three amino acid changes, need to be confirmed in genomic DNA from several animals to confirm that they are genuine and not sequence errors. Once confirmed each putative causal variant can then be tested for an association with the variance in litter size and prenatal survival levels observed between pig breeds.

Chapter Six



6. TESTING CANDIDATE CASUAL SEQUENCE VARIANTS IN SPP1 GENE

6.1. Introduction

Several sequence differences were identified between Meishan and Large White copies of the positional and physiological candidate gene, *SPP1* (chapter 5). Before considering whether these sequence differences are associated with variation in prenatal survival levels and litter size, it was necessary to confirm that the sequence differences were genuine polymorphisms rather than sequence errors or artefacts. The variants that were investigated were the six exonic SNPs, including three that result in changes in the predicted amino acid sequence and the presence or absence of a SINE in intron six. None of the sequence variants occurred within recognised regulatory sequence motifs in the promoter or elsewhere in the gene.

A large data set of litter size records over several parities for around 4000 sows was available from the Pig Improvement Company (PIC). The advantage of these data is that there has been more opportunities to accumulate useful recombination events within these breeding populations than within the three generation experimental cross and this gives a much greater resolution to detect an association between specific marker alleles and performance (see Figure 1-1). If an allele at a specific locus was found to contribute to a difference in litter size within a commercial population, then it would be possible to utilise this information within a marker-assisted selective (MAS) breeding program.

A large number of animals need to be genotyped for the single bialleic SNP of interest, in order to carry out an association analysis with sufficient statistical power to detect a correlation between phenotype and genotype at a single locus. There are many techniques for SNP typing, the system I used involved microarray hybridisation chips.

SNPs are the major contributors to genetic variation and account for around 80 % of all known polymorphisms in the human genome. They are of particular interest as some of those located within genes will directly influence protein structure or gene expression levels, therefore allowing a much clearer understanding of complex traits such as disease resistance (Sayers et al., 2000). Studies of SNPs are also important in the upcoming field of pharmacogenomics (tailored genetic medicine). Fahrenkrug *et al.* (2002) have recently reported the first large-scale assessment of SNP frequency and distribution in the porcine genome.

The aim of this chapter was to first confirm all seven variants in genomic DNA from several pigs within the Meishan x Large White Roslin pedigree cross (outlined in section 2.2.1.); and secondly to investigate whether variation at any of these loci contributes to the differences in prenatal survival and subsequent litter size between the Large White and Meishan breeds. Finally an association analysis was used to test for a correlation between the marker genotypes and their litter in several thousand sows, from commercial breed lines. This was done to investigate whether the alleles associated with a difference in litter size at the QTL, detected in the experimental crosses, were segregating within commercial populations.

The SNP of most interest was SNP7.2 in exon 7, which encodes the amino acid proline from the Large White copy of the sequence and the amino acid serine from the Meishan copy of the sequence. The main control of *SPP1* expression is by post-translational modification, in particular phosphorylation, and this serine residue lies within a casein kinase II phosphorylation site, whereas the proline disrupts this site.

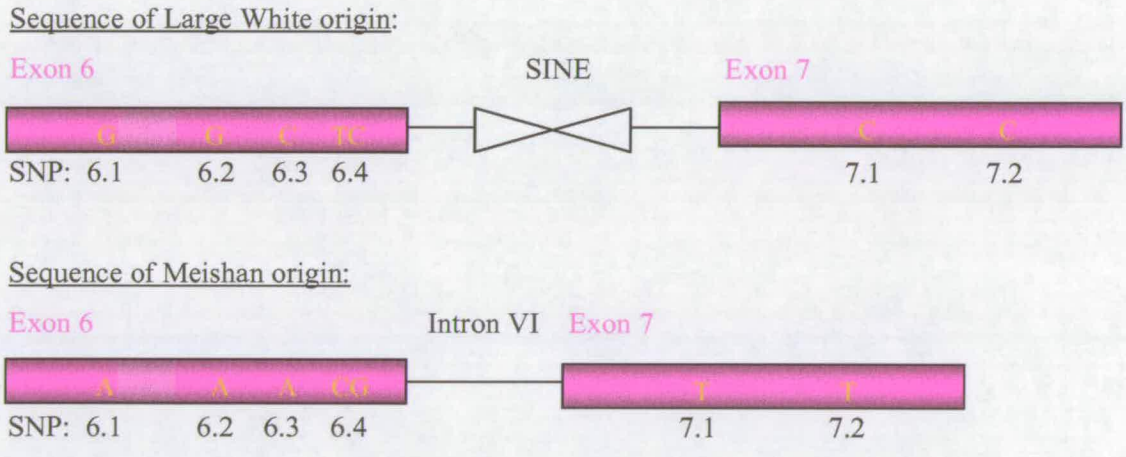
6.2. Materials and methods

(Protocols for all solutions mentioned are detailed in appendix I)

6.2.1. Confirmation of sequence variants in genomic DNA

The sequence variants identified between copies of the *SPP1* gene from Meishan and Large White origin, were tested on genomic DNA samples from individuals in the Roslin Large White x Meishan pedigrees (see section 2.2.1) and on DNA from the Large White x Meishan F1 boar used to sequence *SPP1* and his parents. By typing the pure bred founders the distribution of alleles in these individuals could be assessed and the sequence variants confirmed to be genuine by following their segregation within families.

Initially the three SNPs, which encoded non-synonymous amino acid changes and the presence or absence of the SINE were tested. Figure 6-1 shows the location of all six exonic SNPs and the SINE within *SPP1* for the two sequences (exons 1, 2, 3, 4 and 5 contained no variants in the sequence between the two breeds).



SNPs that encode a non-synonymous amino acid change:
SNP6.1 – (Large White origin) **GCT** = Alanine, (Meishan origin) **ACT** = Threonine
SNP6.4 – (Large White origin) **GTC** = Valine, (Meishan origin) **GCG** = Alanine
SNP7.2 – (Large White origin) **CCC** = Proline, (Meishan origin) **TCC** = Serine

Figure 6-1 The identity of the six SNPs found within the exons of *SPP1* and the position of the SINE for sequences of Large White and Meishan origin.

The methods used to type each of the variants were different and are described in section 6.3. Briefly, the presence or absence of the SINE was tested as a PCR length variant using agarose gel electrophoresis for all 342 animals across the three generations of the QTL pedigrees. SNP 6.1 (encoding alanine or threonine) was typed as a restriction fragment length polymorphism (PCR-RFLP) also over all three generations of the crosses. SNP 7.2 (encoding serine or proline) was typed by Zhihua Jiang (University of Guelph) for the QTL analysis (Chapter 2), as the BI-PASA marker "*SPP1-5*". Finally, SNP 6.4 (encoding valine or alanine) was typed over the eight Fo purebred individuals for the QTL1 population only, by sequencing across the SNP using an ABI 3100 sequencer.

6.2.1.1. DNA Sequencing (ABI 3100 sequencer)

The methodology for the 3100 Genetic Analyser is similar to the 373 (section 5.2.5) and the differences in the technique are outlined. The ABI PRISM® BigDye™ terminator v3.0 cycle sequencing ready reaction kit (Applied Biosystems, Warrington, UK) was used for sequencing sample preparation.

Sequencing reactions were carried out to a total volume of 10 µl; 4 µl ready reaction mix was added to 1 µl of DNA and 1.6 pmol of primer. The following program was used for primer extension and dye termination; 25 cycles of a rapid thermal ramp (1° C/sec) to 96° C, 96° C for 10 sec, a rapid thermal ramp to 50° C, 50° C for 5 sec, a rapid thermal ramp to 60° C, and then 60° C for 4 mins followed by a final rapid thermal ramp to 18° C. The samples were stored at 4° C prior to precipitation. A PTC-225 DNA engine tetrad (MJ Research, Inc, Braintree, UK) was used for the primer extension reaction.

6.2.1.1.1 Ethanol/Sodium Acetate precipitation

A 40 µl mix was added to each sample, consisting of 1.5 µl sodium acetate (NaOAc), 31.25 µl of 95 % ethanol (EtOH) and 7.25 µl double distilled water. The 96-well skirted plate (ABgene, Epsom, UK) was sealed and mixed by inverting a few times. The plate was then incubated at room temperature for a minimum of 1 hour in the dark; this allowed the precipitation of the extension products. It was then spun in a 5804 bench-top centrifuge with a plate adapter (A-2-MTP) (Eppendorf AG,

Hamburg, Germany) at 3000-x *g* for 30 minutes. The supernatant from each well was then discarded by inverting the plate onto a paper towel.

The pellets were then rinsed by adding 75 µl of 70 % ethanol to each well and the plates mixed by inversion and spun at 3000-x *g* for 10 minutes. This wash was then discarded by inverting the plate onto a paper towel and the inverted plate, with the towel under it, spun at 50-x *g* for 1 minute. Unlike the precipitation method of the 373 sequencing, air-drying of the pellet was not required, and the sample in each well was immediately re-suspended and denatured in 10 µl Hi-Di formamide. The re-suspended samples were then transferred to an ABI 96-well plate and the plate loaded onto the 3100 Genetic Analyser. In contrast to slab-gel electrophoresis used with the 373, the 3100 Genetic Analyser uses capillary electrophoresis. Instead of a solid gel, a linear flowable polymer called 3100 POP-6™ polymer is pumped through capillaries and the negatively charged DNA samples are loaded into the 16 capillaries by electrokinetic injection.

6.2.2. Allele association analysis for candidate causal SNPs

It is possible to utilise microarray technology to genotype individual SNPs or bialleic markers over many hundreds of animals. These genotypes can then be combined with the phenotypic records for each animal and an allele association analysis carried out to determine whether variation at the SNP is correlated with improved performance in the trait, in this case litter size.

6.2.2.1. Sows available with phenotypic records

In order to investigate the SNPs of interest within *SPP1,* DNA samples from 4017 sows from commercial breeding populations were obtained from the Pig Improvement Company (PIC). This study allows an independent investigation of the putative causal variants identified within the *SPP1* gene, to see whether the association between the region on SSC8 around this gene and litter size seen in the structured Meishan x Large White crosses, is also present in commercial pig populations. Records of the number of piglets born alive were available for these

sows over several parities. All 4017 sows had a first parity record and progressively smaller subsets of animals had records for the second, third, fourth and fifth parities. The animals were from six different breed origins (Table 6-1).

Table 6-1 The breed origin of each line and the number of animals in each line group

Line identity	Breed origin	Number of animals in line
Line C	Landrace	1089
Line D	Large White	1046
Line E	Duroc	544
Line F	Duroc	335
Line G	Mixed synthetic selected for high lean and growth rate	464
Line H	Synthetic 50 % Large White 50 % Meishan	539

6.2.2.2. SNP genotyping using DNA microarray technology

Target DNA from all 4017 sows is arrayed on the slide and probed with two oligonucleotides, each containing the alternative allele at the SNP. The oligonucleotide probes are labelled with either a cy3 (green) or a cy5 (red) fluorophore label. Therefore the target DNA of homozygous individuals will fluoresce with a signal, which is visualised as either green or red and the target DNA of heterozygous individuals will fluoresce with a composite signal, which appears orange.

6.2.2.2.1 Preparation of target DNA on microarray slides

The 4017 DNA samples were aliquoted into thermo-Fast® 384 well PCR plates (ABgene, Epsom, UK) by individuals from the Pig Improvement Company (PIC). They added 2 µl of the DNA samples of varying concentrations to the appropriate wells on the plates and then dried down the samples at 37° C for two hours and stored the plates at -20° C. In addition at least one of the wells on every plate was left empty, to act as a negative control.

Primers were designed that flanked the SNP of interest and would give a relatively short amplicon of less than 100 bp, to allow efficient binding to the microarray slide, and reduce non-specific background hybridisation from the detection oligonucleotides. Target DNA fragments were amplified in 6 µl reactions. The standard reaction mix included 3 pmol of each primer (MWG Biotech, Germany), 2.0 mM of each of dATP, dUTP, dGTP and dCTP (Applied Biosystems, Warrington, UK), 2.5 mM $MgCl_2$ in 1x PCR buffer (Applied Biosystems, Warrington, UK), 0.5 µg creosol red indicator dye (Aldrich Chemical Company, WI, USA), 0.05 U Uracil DNA glycosylase and 0.25 U AmpliTaq gold (Applied Biosystems, Warrington, UK). Uracil DNA glycosylase (UDG) was added to the PCR mix to eliminate any contamination by terminating amplification from any non-specific template in the reaction mix. This enzyme is activated during the first PCR cycle at 50° C and is then inactivated during the denaturation step at 94° C. The GeneAmp® PCR system 9700 (Perkin Elmer™, Foster City, CA, USA) was used with a touchdown PCR program for target DNA amplification:

Activation of UDG:	50° C 2 min
Denaturation cycle:	94° C 12 min
12x cycles:	94° C 30 sec
	65° C 30 sec* Decrease 1° C per cycle to 53° C
	72° C 30 sec
23x cycles:	94° C 30 sec
	52° C 30 sec
	72° C 30 sec
Extension:	72° C 7 min

2 µl of 4x salt buffer was added to each 6 µl PCR product to create a spotting solution. Due to the high salt content and alkaline pH of the spotting solution, when it is added to the PCR mix it turns the creosol red purple. This colour change was used as an indicator to ensure spotting buffer had been added to every well of the PCR plate. The PCR plates were then spun at 1000 x *g* for 30 seconds (4236A centrifuge (CamLab)). The MicroGrid II (compact) robot (BioRobotics, Cambridge,

UK) was used to prepare the microarray slides. The PCR products for each DNA sample were spotted in triplicate onto poly-lysine slides (BDH, Poole, UK), using either a 16 pin or 48 pin tool, depending on the number of samples to be added to the slide. This poly-lysine coating is positively charged and therefore binds the negatively charged DNA molecules. The pin tools were always cleaned by sonication prior to use. The spot size used was 160 μm and the size and number of arrays on the slide was calculated from the number of DNA samples to be spotted.

The region around the array grid was marked on the glass slide to allow the positioning of the cover slip, once the salt from the spotting solution had been removed during the denaturing step. The slides were then baked at 80° C for 2 hours on a hotplate (model 1201, Jenway, Dunmow, Essex, UK) to immobilise the DNA samples to the glass slide.

The array pattern for the slides printed was then combined with a .CSV file containing the identity of each sample and its location on each PCR plate, to produce a .gal file containing the identity of the DNA samples at every location on the arrays.

6.2.2.2.2 Denaturation of arrays and hybridisation of DNA probes

Prior to hybridisation with DNA probes, the arrays were denatured for 1 minute in boiling water and then rinsed in 100 % ethanol, to separate double stranded PCR amplicons into single stranded DNA. The slides were then immediately spun dry at 1000 x *g* for 1 minute (4236A centrifuge (CamLab)) to prevent dust accumulating on the wet slides. All slides were stored in an airtight box to keep them clean.

For each SNP being tested, individual slides were hybridised with the two fluorescent probes representing the two alternative alleles. These fluorophores are bleached if exposed to light and were therefore stored in the dark at all times. A 60 μl hybridisation mix consisted of 3 pmol of each fluorescent probe (MWG Biotech AG, Ebersberg, Germany), 31.2 μg salmon testes DNA (Sigma-Aldrich Company Ltd, Dorset, UK), 15 μl of 20x SSC and 3 μl of 10 % SDS (diluted from 20 % (w/v) sodium dodecyl sulfate). The salmon testes DNA acts as a DNA blocker

and prevents the probes binding to non-specific sites. 58 µl of the hybridisation mix was placed onto the inner edge of the array region ensuring that no bubbles were formed and a coverslip (24 x 64mm BDH coverglass (thickness no. 1)) carefully added.

40 µl of 5x SSC was added to both wells of a hybridisation chamber and the slide placed into the chamber and hybridised in a water bath at 34° C for 30 minutes. The slides were then washed for 5 minutes in each of 5x SSC with 0.05 % SDS, 5x SSC and 3x SSC. The slides were rinsed in distilled water and spun for 1 minute at 1000 x *g* (4263A centrifuge (Camlab)).

6.2.2.2.3 Analysis of microarrays and genotype identification

Each microarray slide was scanned using a ScanArray® 4000 MicroArray analysis system scanner (GSI lumonics) and visualised with the compatible ScanArray program at the cy5 wavelength of 635 nm and the cy3 wavelength of 532 nm. Using 100 % laser power and a photo-multiplier tube (PMT) gain of between 80 and 90, the intensity of the image for each fluorophore was optimised without enhancing the background noise. The slides were scanned at a resolution of 10 µM and separate cy5 and cy3 images produced.

Using the program GenePix 3.0, the cy3 and cy5 image files were combined with the "gal" file containing the location and identity of the DNA samples within each array. The grid of DNA sample locations was aligned with the appropriate spots on the array. For the image analysis only the inner area of each spot was considered, in order to include only the regions of most intense colour and to minimise the background. Only the colour of the spot i.e. green, orange or red is of interest for genotyping, not the intensity of the whole spot as for studies of levels of gene expression. Any samples that had failed were highlighted.

The intensity of fluorescence for both the cy3 and cy5 signals for each spot on the slide was exported from GenePix and analysed within Microsoft Excel. Initially a

scatter plot was produced of the mean intensity of the cy5 (F635) and the cy3 (F532) signals over each spot, in order to ensure that three clear regions were seen on the graph corresponding to the three genotype classes. Each sample was spotted on the slide in triplicate, therefore the mean of the average intensity of the cy5 and cy3 signals (F635 and F532) was calculated for the three replicates of each DNA sample.

The intensity of the signals for the negative controls or the samples that had failed represents the background noise on the slide and was used to determine a threshold value, above which the hybridisation of the DNA samples was considered to have been successful. A table was produced of the mean cy5 and cy3 intensity for every animal that had been successfully genotyped.

The $\log_{10}$ ratio and the sum of the two mean intensities were calculated and plotted as a scatter graph. The three genotypic classes were detected as the distinct clusters of spots of varying intensity of cy3 and cy5 fluoresence. The range of $\log_{10}$ ratio values for each genotype class was then used to classify each individual and to exclude any data points that displayed intensities lower than the background threshold.

6.2.2.3. Allele association analysis

6.2.2.3.1 One-way ANOVA

Initially a one-way analysis of variance (ANOVA) was carried out using Minitab™ statistical software (release 13.32) to determine whether there was a significant effect of a sow's genotype at the individual SNP typed, on the number of live piglets she farrowed. The null hypothesis was that the mean number of piglets born alive is not significantly different for sows of varying genotype at the putative causal SNP. The analysis was carried out within single parity records for each sow.

The level of statistical significance (*P* value) was then determined by looking up the critical value of the F statistic variance ratio for a one-tailed test at the appropriate degrees of freedom. If the null hypothesis is true, then the two estimates of variance (mean square) will be similar and the F value will be around 1. Where the *P* value is less then 5 % (0.05) the null hypothesis can be rejected and it can be said that there is

a significant difference in number of live piglets between genotype classes. The 95 % confidence interval for the mean number of live piglets within each genotype class was then calculated as SEM (standard error of the mean) x 1.96, where SEM = pooled SD / $\sqrt{n}$ (SD = standard deviation across genotype classes and n = number of sows).

6.2.2.3.2 REML analysis

A REML (Residual Maximum Likelihood Estimation) analysis was carried out for sows at each parity level using GenStat® for Windows, sixth edition. A linear mixed model of REML was used as this not only tests for an association between sow genotype and litter size, but also accounts for the major factors affecting litter size. These include the fixed effects of farm background, breed origin, number of mating services of the sow and genotype. The identity of the sire of each sow was included as a random effect.

As the six breed groups were reared on two different farms the farm identity was included as a fixed effect, to account for environmental noise, such as differences in management and nutrition. Table 6-2 shows the number of sows having their first parity litter (n= 2940) within each of six breed groups that were raised on each of the two farms. There were a total of 511 sires of the sows, and no one sire was used on both farms. The breed origin refers to the six line groups. The number of services by the boar ranged from one to six, with the majority of sows only mated once; it is unlikely that this parameter will have a large influence on litter size however all possible influential variables should be included in the analysis. The identity of the sire of the sows was included to account for some of the genetic variation in litter size and some of the pedigree relationships between sows.

Table 6-2 Number of sows (first parity) in each of the six breed groups raised on the two farms (A and B). Line C = Landrace, D = Large White, E = Duroc, F = Duroc, G = mixed synthetic, H = 50 % Meishan.

Farm	Breed line					
	C	D	E	F	G	H
A	470	463	250	0	0	338
B	432	416	0	249	322	0

6.3. Results

6.3.1. Confirmation and testing of candidate sequence variants within Roslin Large White x Meishan pedigree cross

6.3.1.1. SINE repeat element

A region of 1.0-1.3 kb, encompassing the 3' end of exon 6, intron 6 and the 5' end of exon 7, was amplified using the primer pair described by Knoll *et al.* (1999). Fragments of different lengths were amplified depending on whether the individual animal possessed the 330 bp SINE within intron 6 of the *SPP1* gene, and this difference in size was detected by agarose gel electrophoresis. All 342 animals from the three Meishan x Large White cross populations were tested in duplicate. The F_1 heterozygous boar used to create the BAC library and his purebred Large White and Meishan parents were also typed. The DNA from the F_1 boar acts as a positive control, as the copy of *SPP1* from Meishan and Large White origin was sequenced from the BAC library (see section 5.3.1) and therefore SNPs typed over this DNA sample should always be heterozygous.

A 25 µl reaction mix was made for each genomic DNA sample. This PCR mix consisted of 7 pmol of each primer (MWG Biotech AG, Ebersberg, Germany), 2.0 mM $MgCl_2$ in 1x PCR buffer (Roche Diagnostics, Mannheim, Germany), 2.0 mM of dTTP, dCTP, dGTP and dATP (Amersham Pharmacia Biotech inc, Little Chalfont, UK), 0.5 U *Taq* DNA polymerase (Roche Diagnostics, Mannheim, Germany) and 50 ng of genomic DNA.

A PTC-225 DNA engine tetrad (MJ Research) was used with a hot start PCR program. After an initial 95° C denaturation step for 2 minutes, the anneal phase consisted of 29 cycles of 95° C for 45 seconds, 65° C for 1 min and 72° C for 2 minutes and finally an extension cycle of 72° C for 7 minutes.

The PCR products were then separated through a 2.5 % Ultra pure DNA grade agarose gel (Bio-Rad laboratories, Hercules, CA, California) in 1x TBE buffer,

stained with 160 µg ethidium bromide and run alongside a 1 kb plus DNA ladder (Invitrogen Life Technologies). For those individuals possessing a copy of the gene with the SINE element, the DNA fragment amplified was around 1.3 kb in size and for those with a copy of the gene without the SINE a fragment of around 1.0 kb was amplified. Bands of both sizes were seen for heterozygous individuals. Figure 6-2 shows an example gel separation for the Fo purebred individuals from the three QTL populations. Also shown is the heterozygous boar used to sequence the *SPP1* gene and his parents.

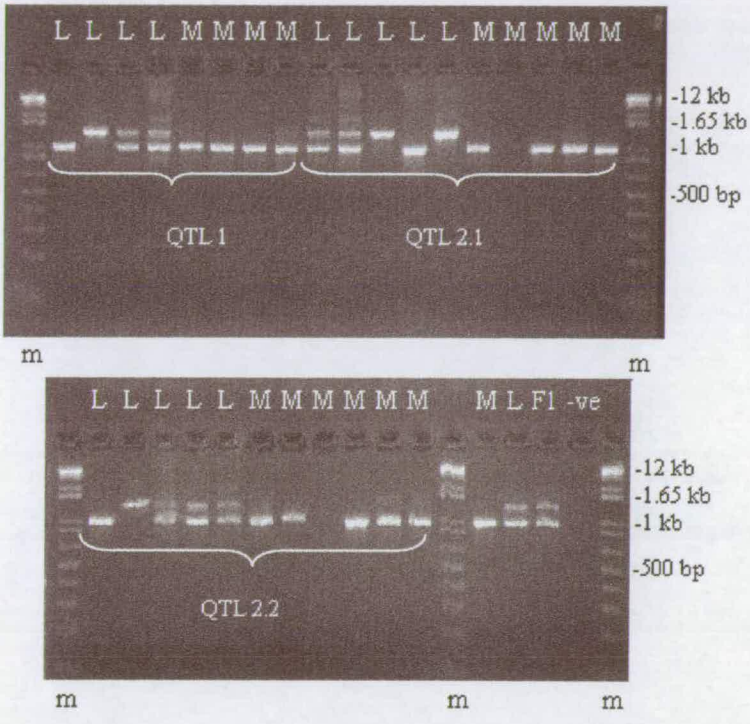


Figure 6-2 Genotyping of Meishan (M) and Large White (L) Fo purebred animals from the three Roslin QTL populations, the F1 boar used to sequence the *SPP1* gene and his Meishan sire and Large White dam for the identification of the presence or absence of the SINE in intron 6. (m= 1 kb plus DNA ladder size standard).

All the purebred Meishan individuals are homozygous for the absence of the SINE repeat. The PCR for two of the DNA samples had been unsuccessful, however the repeat assay showed that both individuals were also homozygous negative. The

allele for the presence of the SINE element was segregating within the Large White individuals; some individuals were homozygous for the repeat element, some were heterozygous and others homozygous negative. In agreement with the *SPP1* gene sequence data, the F1 individual used to create the BAC library was shown to be heterozygous, with the allele containing the SINE element having been inherited from his Large White dam, who was also shown to be a heterozygote.

6.3.1.2. SNP encoding alanine or threonine (SNP6.1)

The base change from guanine to adenine present in exon 6 that encodes an amino acid change from alanine to threonine (see base position 8449 appendix II and residue 110 on Figure 5-13) was typed with a PCR-RFLP (restriction fragment length polymorphism) assay. Webcutter version 2.0 (Maarek et al., 1997) (http://www.firstmarket.com/cutter/cut2.html) was used to identify restriction endonuclease sites, that would allow detection of variation at the SNP locus. For example, the allele containing base G would be cut with the enzyme *Mwo*I and the allele containing the base A would not (Figure 6-3).

```
LW:          GACTCCGAGGAAGCTGATCACGCTGACGACGCTGACCGATCCG
Recognition site(MwoI):          GCNNNNNNNGC
MS:          GACTCCGAGGAAGCTGATCACACTGACGACGCTGACCGATCCG
```

Figure 6-3 *Mwo*I restriction site. The allele with base G at the SNP will be digested and the allele with base A at the SNP will not. (Only single stranded DNA is shown, however the reverse sequence will also be digested).

One of the primer pairs (SPPI5aF and SPPE6R) used for sequencing, flank this restriction site and yield a product of 515 bp (for primer sequences see Table 5-3). However a second *Mwo*I restriction site was discovered to lie around 50 bp downstream of the forward primer. Therefore a new forward primer was designed around 70 bp downstream of this site, 5'-TTAGACCCTGCCAAGCAAGT-3' (SPPE6Fb). The resulting product was 391 bp and where *Mwo*I cut the fragments produced were 121 bp and 270 bp.

All 342 animals from the three Meishan x Large White populations were tested in duplicate. A 20 µl reaction mix was made for each genomic DNA sample. This PCR mix consisted of 5 pmol of each primer (MWG Biotech AG, Ebersberg, Germany), 1.5 mM $MgCl_2$ in 1x PCR buffer (Roche Diagnostics, Mannheim, Germany), 2.0 mM of each dTTP, dCTP, dATP and dGTP (Amersham Pharmacia Biotech inc, Little Chalfont, UK), 0.5 U *Taq* DNA polymerase (Roche Diagnostics, Mannheim, Germany) and 50 ng of genomic DNA.

A PTC-225 DNA engine tetrad (MJ Research) was used. After an initial 95° C denaturation step for 5 minutes, the anneal phase consisted of 35 cycles of 95° C for 45 seconds, 58° C for 1 min and 72° C for 2 minutes and finally an extension cycle of 72° C for 7 minutes. A few random 2 µl samples of the PCR products from each PCR plate were then separated through a 100 ml 2.5 % Ultra pure DNA grade agarose gel in 1x TBE buffer, stained with 40 µg ethidium bromide and run alongside a 100 bp DNA ladder (Invitrogen Life Technologies), in order to confirm that the PCR had been successful and that PCR products of the correct size (391 bp) had been amplified, before carrying out the restriction digest.

The PCR products (10 µl) were digested with 10 µl of 2.5 U of *Mwo*I and 2x Buffer (New England Biolabs) at 60° C for 16 hours. In order to ensure thorough mixing of the enzyme, the PCR plates were spun at 880 x *g* for 1 minute on a 5804 bench-top centrifuge with a plate adaptor (A-2-MTP) (Eppendorf). The resulting restriction fragments (20 µl) were run with a 100 bp DNA ladder (Invitrogen Life Technologies) on a 2.5 % Ultra pure DNA grade agarose gel stained with 160 µg ethidium bromide. Figure 6-4 shows examples of the gel separation for the F_0 purebred individuals from the three QTL populations. Also shown is the F_1 heterozygous boar used to sequence the *SPP1* gene and his parents. All purebred Large White individuals were homozygous for the base G at SNP6.1, encoding alanine. However this SNP is segregating within the Meishan breed and all three genotypes *AA*, *AG* and *GG* were observed. In agreement with the *SPP1* gene sequence data, the F1 individual used to create the BAC library was shown to be heterozygous, and his parents homozygous for alternative alleles.

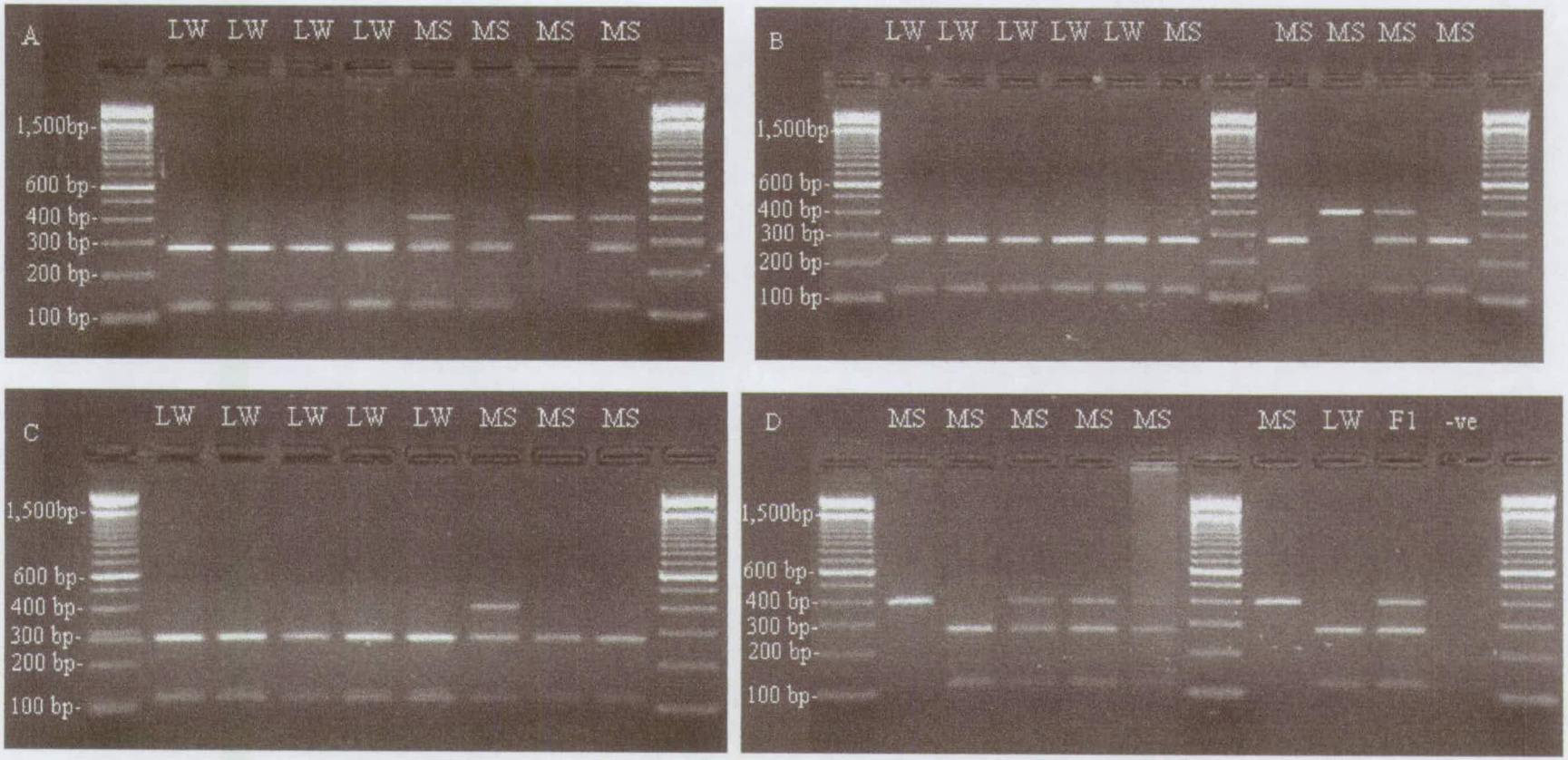


Figure 6-4 PCR-RFLP assays of SNP6.1 from F0 DNA samples from three QTL populations alongside a 100 bp DNA ladder size standard. MS = Meishan pure-bred individual, LW = Large White pure-bred individual. The allele with base G at the SNP is digested to yield 121 bp and 270 bp fragments and the allele with base A at the SNP remains as an undigested 391bp PCR product. (A – QTL1, B- QTL2.1, C- QTL2.2, D- QTL2.2 ctd, F1 boar used to sequence the *SPP1* gene and his parents and negative control).

6.3.1.3. SNPs encoding valine or alanine (SNP6.4)

The second variant of interest in exon 6 includes two adjacent SNPs, with base changes from TC to CG, and causes a change from valine to alanine (see appendix II base position 8597-8598 and residue 159, Figure 5-13). As with SNP6.1, potential RFLP sites were investigated. Unfortunately no suitable restriction sites were found.

The valine to alanine is a conserved change, as both amino acids have non-polar side chains, alanine has one methyl group and valine has three. It seems unlikely that this sequence variant would affect differences in litter size and prenatal survival traits. However it was still necessary to confirm the changes observed from the BAC sequencing in genomic DNA, in order to be certain that the sequence variation was genuine. Therefore the region around this variant was sequenced across the eight QTL1 F_0 purebred individuals from both breeds, the F_1 boar used to create the BAC library and his purebred Large White and Meishan parents.

Primers were designed that flanked the two base pair variant of interest in order to PCR this region from genomic DNA. Forward and reverse internal primers were then used to sequence across the variant. Figure 6-5 shows the design of this experiment. The primers were designed from sequence for which there was no evidence of sequence variation between the two breeds.

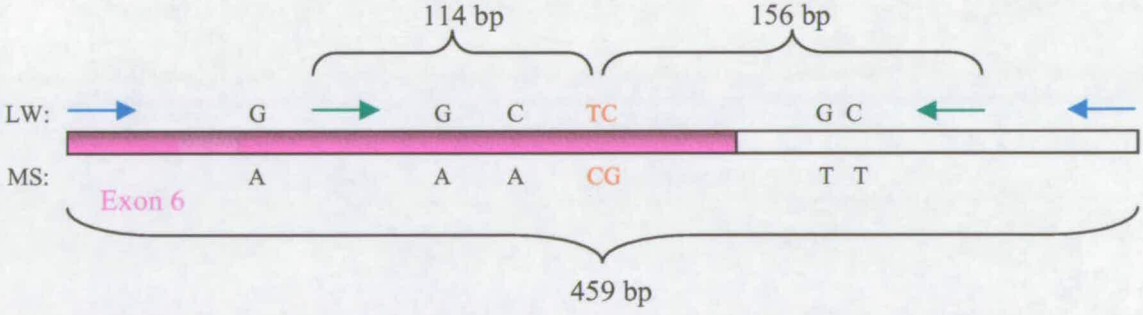


Figure 6-5 Forward and reverse primers (shown as blue arrows) were used to amplify 459 bp PCR product around the two SNPs in exon 6 (shown in red). Other SNPs in the region are shown in black. Internal primers (shown in green) were used to sequence in a forward and reverse direction across the SNPs.

The expand high fidelity PCR system, which includes a proofreading polymerase, was used with the forward (5'-GAGCAAACAGACGATGTGGA-3') and reverse (5'-TGGAAAAACTGCCTCGTGAT-3') primers to amplify the 459 bp region. This system allows efficient amplification of DNA fragments up to 5 kb from genomic DNA. Due to the 3'-5' exonuclease proofreading activity of the enzyme, this system results in a 3-fold increased fidelity of DNA synthesis (8.5×10^{-6} error rate) compared to *Taq* DNA polymerase (2.6×10^{-5} error rate). This proofreading capability therefore reduces the chance that mis-match bases will be introduced during the PCR.

When the whole of *SPP1* was sequenced (chapter 5) a proofreading enzyme was not required as several PCR products were combined from different PCR reactions. The chance of sequence error in a single PCR reaction is actually reasonably low, in particular because any misincorporation errors tend to promote chain termination and the amplification of defective molecules is restricted (Innes and Gelfand, 1990). However, for the confirmation of any putative sequence variants in this single PCR reaction, the use of this enzyme was desirable simply to further reduce the chance of error.

For the PCR reaction two 10 µl master mixes were made containing sufficient reagents for all 11 DNA samples. The first consisted of 3 pmol of each primer (MWG Biotech AG, Ebersberg, Germany), 2.0 mM of each dNTP (Amersham Pharmacia Biotech inc, Little Chalfont, UK), and 50 ng of genomic DNA (added separately to PCR plate). The second consisted of 1.5 mM $MgCl_2$ in 1x Expand HF buffer and 2.6 U Expand HF PCR system enzyme mix (Roche Diagnostics, Mannheim, Germany).

The reagents from the two master mixes were added together just prior to the PCR, therefore avoiding the need for a hot start and also preventing the interaction of the enzyme mix with primers or template in the absence of dNTPs. This could lead to partial degradation of the primer and template through the 3'-5' exonuclease activity of the proofreading enzyme. 96-well skirted PCR plates were used (ABgene).

The PCR was carried out on a PTC-225 DNA engine tetrad (MJ Research). After an initial 94° C denaturation step for 2 minutes, the anneal phase consisted of 30 cycles of 94° C for 15 seconds, 60° C for 30 seconds and 72° C for 45 seconds and finally an extension cycle of 72° C for 7 minutes.

The PCR products (20 µl) were then run with a 100 bp DNA ladder (Invitrogen Life Technologies) on a 1 % Ultra pure DNA grade agarose gel stained with 40 µg ethidium bromide, to confirm that sufficient 459 bp product had been amplified for each DNA sample (Figure 6-6). Insufficient PCR product for DNA sample 8 was produced and the PCR and gel purification were repeated, using freshly diluted DNA from the stocks. The repeat is shown alongside the other samples.

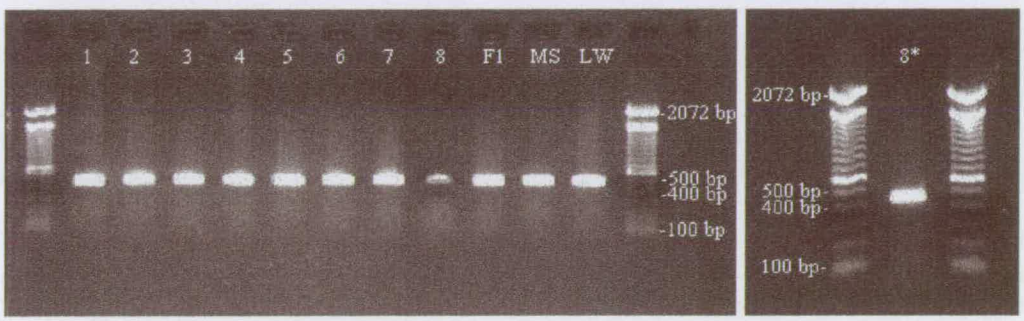


Figure 6-6 Amplification of 459 bp fragment around SNP6.4. DNA samples 1-4 were Large White purebred individuals and samples 5-8 were Meishan. Sample 8 was repeated using freshly diluted DNA from a concentrated stock sample and the repeat is shown as 8*. F_1 = heterozygous boar used to sequence the *SPP1* gene, MS and LW = Meishan and Large White parents of F_1 boar.

The PCR products were then gel purified and the resulting concentration of eluted DNA determined. Figure 6-7 shows the estimated concentration of each of the DNA samples. For sequencing using the 3100 Genetic Analyser 3-10 ng of purified PCR product (200-500 bp) was required and it can be seen that all of the purified DNA have a concentration of around 5-10 ngμl^{-1}. Therefore 1 µl of each DNA sample was included in the sequencing reaction, with each of the internal sequencing primers. The eleven DNA samples were sequenced in the forward and reverse direction using the ABI 3100 Genetic Analyser. The forward primer sequence was 5'-

ACTCCGATGAATCCGATGAG-3' and the reverse was 5'-CAAGTAGTCATGCTGGCTGAA-3'.

The sequence traces were viewed using Chromas (available from Technelysium Pty Ltd, http://www.technelysium.com.au/chromas.html) and the overall quality of sequence was seen to be high. All of the Large White individuals were homozygous for the *TC* allele, one Meishan individual was clearly heterozygous and another appeared to be heterozygous, although the intensity of amplification of one of the alleles was much lower than the other. Another Meishan was homozygous for the *CG* allele and one was homozygous for the *TC* allele. Example sequence traces for each genotype class are shown in Figure 6-8. The pattern of allele distribution for each of the eight QTL1 animals was found to be identical for SNP6.1 and SNP6.4, demonstrating that these two loci are in linkage disequilibrium in these purebred individuals. In agreement with the *SPP1* gene sequence data, the F1 individual used to create the BAC library was shown to be heterozygous and his parents homozygous for alternative alleles.

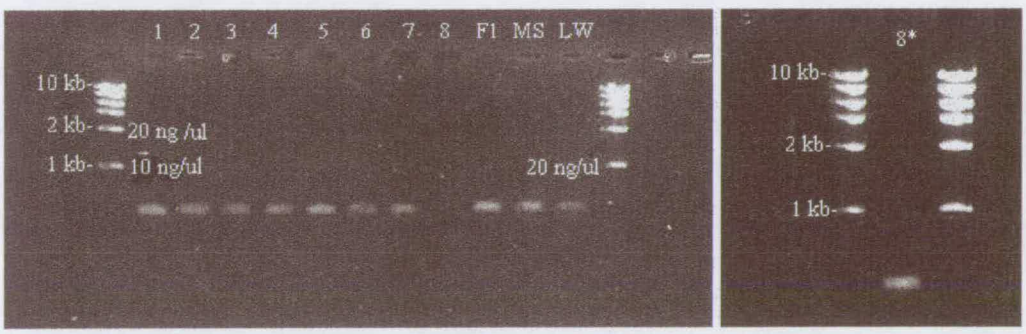
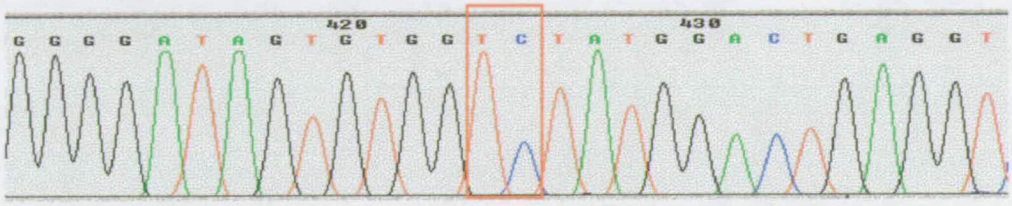
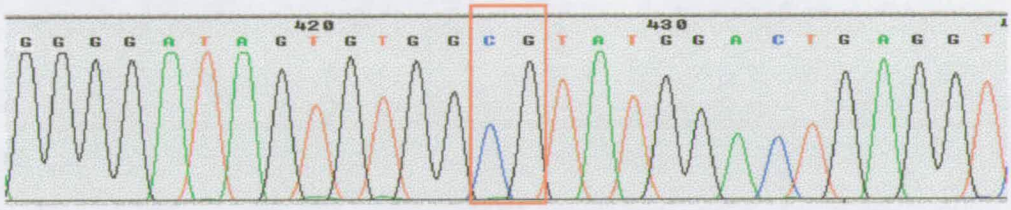


Figure 6-7 Determination of gel purified DNA concentration. DNA samples 1-4 were Large White individuals and samples 5-8 were Meishan. Sample 8 was repeated using freshly diluted DNA from a concentrated stock sample and the repeat is shown as 8*. F_1 = heterozygous boar used to sequence the *SPP1* gene, MS and LW = Meishan and Large White parents of F_1 boar.

Homozygous for Large White allele (TC):



Homozygous for Meishan allele (CG):



Heterozygous (C/T G/C):

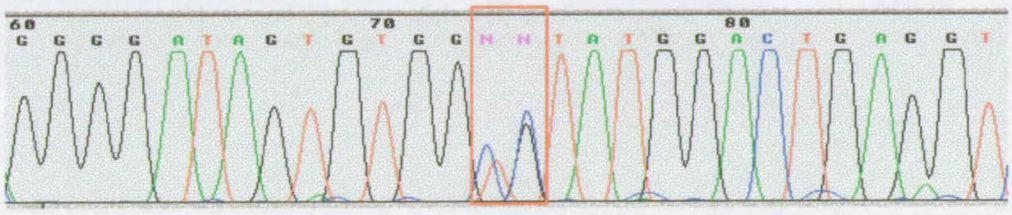


Figure 6-8 Example ABI sequence traces from Chromas of a homozygous individual for the *TC* allele of Large White origin, a homozygous individual for the *CG* allele of Meishan origin and a heterozygous animal for SNP6.4.

6.3.1.4. SNP encoding proline or serine (SNP7.2)

The third SNP encoding a non-synonymous amino acid change is in exon 7 (base position 9919 to 9921 appendix II and residue 262 (Figure 5-13). Sequence traces from the Gap4 output of the BAC DNA sequencing illustrate this SNP (Figure 6-9). Base C at the SNP in the Large White copy of *SPP1* encodes a proline and base T in the Meishan copy of the gene encodes a serine. A Bi-PASA marker "*SPP1-5*" had previously been used by Zhihua Jiang (University of Guelph) to type this locus within the three QTL populations and it was found to be fixed for the alternative alleles in the two breeds. All of the thirteen F0 Large White purebreds were homozygous for base C encoding the amino acid proline and all seventeen F0 Meishan purebreds were homozygous for base T encoding the amino acid serine. Therefore this marker is fully informative within the reproductive QTL population.

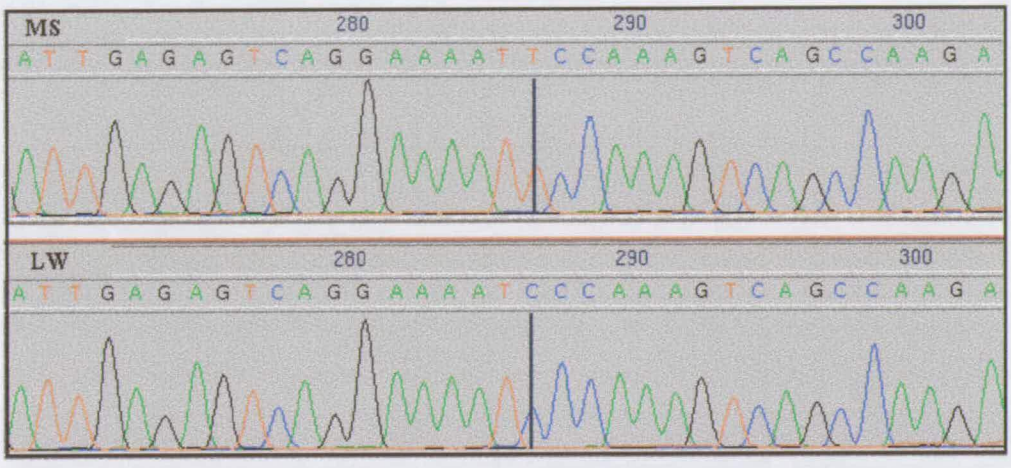


Figure 6-9 Sequence traces from Gap4 output, demonstrating SNP7.2 in exon 7. The allele from the Meishan (MS) copy of the gene contains base T and the allele from the Large White (LW) copy contains base C.

6.3.1.5. Test of variants for within breed marker associated variation at the prenatal survival / litter size QTL

The bialleic loci for the first SNP in exon 6 (SNP6.1 on Figure 6-1) and for the presence/absence of the SINE had both been typed over the three QTL populations and both were shown not to be fixed in the founder breeds. Therefore, as described in section 2.2.6., these gene-associated polymorphisms, which are not fixed for alternative alleles in the founder breeds, were used to test for within breed trait variation at the *SPP1* marker.

SNP6.1 was shown to be segregating in the Meishan breed and fixed in the Large White breed. In contrast, the allele containing the SINE repeat was seen to be segregating within the Large White breed and was not present at all in the Meishan breed. The genotypes for SNP6.1 and for the presence/absence of the SINE were combined with the data for the 20 markers used in the QTL scan, to construct a revised linkage map. This was done in the same way as outlined in section 2.2.4. The resulting map was 139 cM in length, compared to the original map of 139.3 cM. All the markers were in the same order as the previous map, with all of the four

markers within the *SPP1* gene (SNPex6a, SNPex7, SINE element and the microsatellite) mapping to the same position.

There was seen to be no variation in the prenatal survival and litter size traits associated with both of these gene markers over and above any due to the QTL. For both of the variants, there was seen to be no significant difference (P <0.01) between the full model of the normal QTL scan and the model where the marker genotypes for the SINE element and SNP6.1 were fitted as fixed effects. In agreement with this finding, there was no significant variation in the effect of the prenatal survival and litter size traits, between the different classes of genotypes.

6.3.1.6. Use of microarray technology to genotype all 6 exonic SNPs

Out of the six SNPs found within exonic regions of *SPP1*, only the three SNPs that encoded an amino acid change had been confirmed in genomic DNA (see section 6.3.1). I decided that it would be worthwhile to use the microarray genotyping method to verify all six SNPs found in exons 6 and 7 of *SPP1* (Figure 6-1). This would also allow the reliability of the microarray genotyping method to be tested by comparing the genotypes produced with those of the three SNPs that had already been typed. DNA from the 91 individuals in the Roslin Large White x Meishan QTL1 population (see section 2.2.1) and from the F1 boar used to sequence *SPP1* and his parents were genotyped for all six SNPs. Four primer pairs were used to amplify the regions around the SNPs, producing amplicons no larger than 100 bp (Table 6-3).

PCR was carried out for all 94 individuals and two negative control samples containing distilled water, for each of the four regions of exons 6 and 7. 1 µl of each DNA sample and the controls were arranged into a 384 well plate and the appropriate PCR mix added. The program used for PCR was the same as outlined in section 6.2.2.2.1, except only 21 cycles of the second stage of annealing were used. This was because the concentration of the DNA samples was known to be 50 $ng\mu l^{-1}$, which was believed to be greater than that of the PIC litter size collection DNA.

Table 6-3 Sequences of primers used to amplify four regions around the six SNPS in exons 6 and 7 and the size of the resulting amplicons.

Primer sequences (F=forward, R=reverse)	Amplicon size	SNPs within the amplicon
F- 5' GACACGGACTCCGAGGAAG 3' R- 5' CATCGGAGTGATGAGACTCG 3'	70 bp	SNP6.1
F- 5' CAGCAACCGACGTCACTC 3' R- 5' GGAACTTCTTAGATTTTGACCTCA 3'	99 bp	SNP6.2, 6.3 and 6.4
F- 5' AAAGTGAGGAGACGGATGGT 3' R- 5' AAGCCACGTGCAGGC 3'	64 bp	SNP7.1
F- 5' TGATGGGAGCAATGAGCAT 3' R- 5' GGCTGTGGAATTCTTGGCT 3'	74 bp	SNP7.2

Microarray slides were printed of the 384 samples in triplicate using the 16-pin spotting tool. All slides were denatured to expose single stranded DNA and individual slides hybridised with cy3 and cy5 probes for each of the SNPs. The sequences of the two probes, for each of the six SNPs, are shown in Table 6-4.

Table 6-4 Sequences of the cy3 and cy5 probes for each of the two alleles at the six SNPs in exons 6 and 7

SNP identity	SNP allele (breed origin)	Cy3 probe (green fluorescence)	Cy5 probe (red fluorescence)
6.1	G (LW) A (MS)	5'TGATCACGCTGACGA 3'	5'TGATCACACTGACGA 3'
6.2	G (LW) A (MS)	5'TCCCCACGGAGACC 3'	5'TCCCCACAGGAGACC 3'
6.3	C (LW) A (MS)	5'ATGGCCGCGGGGATA 3'	5'ATGGCCGAGGGGATA 3'
6.4	TC (LW) CG (MS)	5'AGTGTGGTCTATGGACT 3'	5'AGTGTGGCGTATGGACT 3'
7.1	C (LW) T (MS)	5'CCATCCTCGTTGCCC 3'	5'CCATCCTTGTTGCCC 3'
7.2	C (LW) T (MS)	5'GGAAAATCCCAAAGT 3'	5'GGAAAATTCCAAAGT 3'

Due to the relatively small number of DNA samples on the slide, 16 arrays were printed and only 20 µl of hybridisation mix was required to cover the arrays. The hybridisation mix was placed onto the inner edge of the array region and a smaller

coverslip added (22 x 22mm BDH coverglass (thickness no. 1)). Each slide was then hybridised at 34° C for 30 minutes. The slides for all 6 SNPs were analysed and the genotypes for the 94 individuals determined.

The genotypes were calculated with a reasonably high level of success. For SNP6.1 genotypes were acquired for 91 out of the 95 animals and there was one mismatch with the genotypes produced using the PCR-RFLP method (section 6.3.1.2). Also the F1 boar was seen to be heterozygous, having inherited an allele from each of his respective homozygous parents. For SNP6.2 90 out of the 95 animals were successfully genotyped. As this SNP had not previously been typed, the pattern of inheritance across all the families in the QTL1 population was checked and the genotype for one individual did not correspond to alleles that could have been inherited from the parents. As expected the F1 boar was heterozygous and both his parents were homozygous for alternative alleles.

For SNP6.3 86 out the 95 animals were successfully genotyped. When the pattern of allele inheritance across families were checked, there was seen to be three individuals whose alleles at this locus could not have been inherited from their parents. The F1 boar was heterozygous and the Large White parent homozygous for one allele, the genotype of the Meishan parent was unknown. For SNP 6.4 73 animals were successfully genotyped and for the five F0 animals with genotypes, the typings were the same as seen with the sequencing method (section 6.3.1.3). In addition three animals had genotypes that could not have been inherited from their parents. The genotypes of the F1 boar and his parents were not determined.

The genotyping of SNP7.1 did not work at all as the intensity of fluorescence of the spots was too low. It is possible that the PCR failed when amplifying the region around the SNP and would therefore need to be repeated to confirm if the SNP is genuine or not. Finally the genotyping of SNP7.2 was less successful, with only 31 out of the 95 animals typed. Those animals that were typed had the same genotypes as when this SNP was typed as a BI-PASA marker for the QTL analysis (section

6.3.1.4). Again the F1 boar was seen to be heterozygous, having inherited alternative alleles from the homozygous parents.

The eight QTL1 F0 Meishan and Large White individuals have now been genotyped for nine loci within *SPP1*; the microsatellite repeat 5' of exon 1, a SNP within intron 1 (base position 2770 appendix II), the six SNPs in exons 6 and 7 and the SINE in intron 6.

Figure 6-10 shows the genotypes (where known) at each of these loci for the F0 animals. There is only one Meishan (88-0433) and one Large White individual (88-0521) that are homozygous for one of the alternative alleles at all of the nine loci. Interestingly SNP7.2 was the only locus that was seen to be completely fixed for alternative alleles in the purebreds of different breed from all three QTL populations. The SNP in intron 1 was fixed in the QTL1 F0 animals (Figure 6-10), however as Figure 5-7 shows two Meishan F0 animals from the QTL 2.1 and 2.2 populations were seen to be heterozygous at this locus (all the other Meishans were homozygous for allele *G*) and all the Large White F0 animals from all three QTL populations were homozygous for the *T* allele.

6.3.2. Genotyping candidate variants in SPP1 gene to test for association with litter size traits in large independent commercial pig populations

Of the variants identified within *SPP1,* SNP7.2 was the most interesting. Not only was the locus fixed for alternative alleles within the Meishan and Large White QTL pedigree individuals it is also a candidate causal variant. The expression of *SPP1* is mainly controlled by post-translational phosphorylation and base T at SNP7.2 encodes a serine, which lies within a casein kinase II recognition site, whereas base C encodes a proline and therefore no phosphorylation site (see section 5.4). Therefore this SNP was the first to be genotyped across the PIC litter size collection of 4017 animals.

	Microsatellite	Intron 1 SNP	SNP6.1	SNP6.2	SNP6.3	SNP6.4	SINE	SNP7.1	SNP7.2
Meishan: (88-0153)	3	G	G	G	C	TC	Absent	?	T
	4	G	A	A	A	CG	Absent	?	T
Meishan: (88-0433)	4	G	A	A	A	CG	Absent	?	T
	4	G	A	A	A	CG	Absent	?	T
Meishan: (88-0497)	3	G	G	G	C	TC	Absent	?	T
	3	G	G	G	C	TC	Absent	?	T
Meishan: (88-1104)	3	G	G	G	C	TC	Absent	?	T
	4	G	A	A	A	CG	Absent	?	T
Large White: (88-0227)	2	T	G	G	C	TC	Absent	?	C
	3	T	G	G	C	TC	Absent	?	C
Large White: (88-0521)	1	T	G	G	C	TC	Present	?	C
	1	T	G	G	C	TC	Present	?	C
Large White: (88-0833)	2	T	G	G	?	TC	Present	?	C
	3	T	G	G	?	TC	Absent	?	C
Large White: (88-0856)	2	T	G	G	?	TC	Present	?	C
	2	T	G	G	?	TC	Absent	?	C

Figure 6-10 Genotypes at nine loci within *SPP1* for the eight QTL1 F0 Meishan and Large White animals. Alleles highlighted in blue were identified from the Large White copy of the gene and alleles highlighted in orange identified from the Meishan copy of the gene. For the polymorphic microsatellite repeat, allele "3" was present in both breeds and is therefore highlighted in green. Alleles "1" and "2" were unique to the Large White breed and allele "4" was unique to the Meishan breed. A question mark represents an unknown genotype.

The amplicons encompassing SNP7.2 for all 4017 DNA samples were spotted in triplicate onto three separate batches of slides. The first set of 768 amplicons were spotted with a 16 pin tool creating a grid of 16 arrays and the other two sets of 1820 and 1429 amplicons were spotted with a 48 pin tool creating grids of 48 arrays. One of the slides printed with the smaller grid was hybridised with 20 µl hybridisation mix and the 22 mm x 22 mm coverslip added and the two slides with the larger grids were hybridised with 60 µl hybridisation mix and the 24 mm x 64 mm cover slip added.

When the slide with 768 amplicons was analysed, the quality of the results was very high. Figure 6-11 shows the arrangement of the triplicate spots within each array. The plot of the $\log_{10}$ ratio against the sum of the mean intensity of cy5 (F635) and cy3 (F532) signals for each sample, showed a clear separation of the three genotype classes and each individual could be classified with confidence (Figure 6-12). There were clearly fewer animals of genotype *TT* than animals of genotype *CC*.

The quality of results for the other two slides was not as high and it was more difficult to clearly differentiate the genotypic classes. In addition all three slides were seen to have slightly elongated "comet tails" on the spots, where the DNA spot has pulled away from the slide during denaturing. Therefore a series of experiments were carried with the aim of determining the underlying problem and improving the quality of the image. These are described in Appendix III.

Of the conditions tested, the slides hybridised with fresh aliquots of probes at a concentration of 3 µM were of the highest quality and therefore the data from these slides was analysed to ascertain the genotypes of the remaining 3249 animals. The spots for several samples were still too faint relative to the background noise and were excluded from the analysis. Only those animals for which genotype classification had been determined with confidence were included. As a consequence only 2940 animals out of the 4017 with phenotypic data were treated as successfully genotyped. This number is still sufficiently large to give enough statistical power for an allele association analysis.

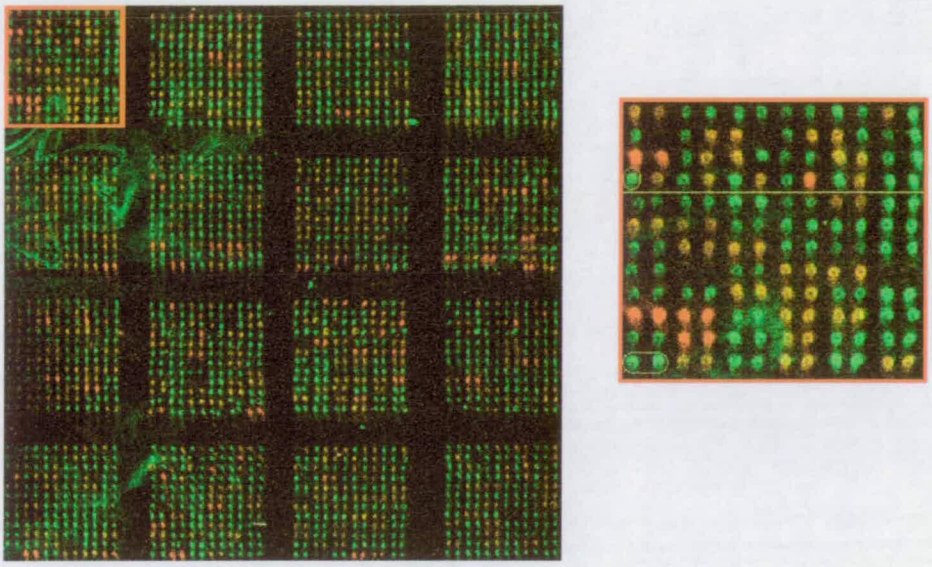


Figure 6-11 The arrangement of the triplicate spots (one set for an individual DNA sample is circled in yellow) within each array on the 4x4 grid.

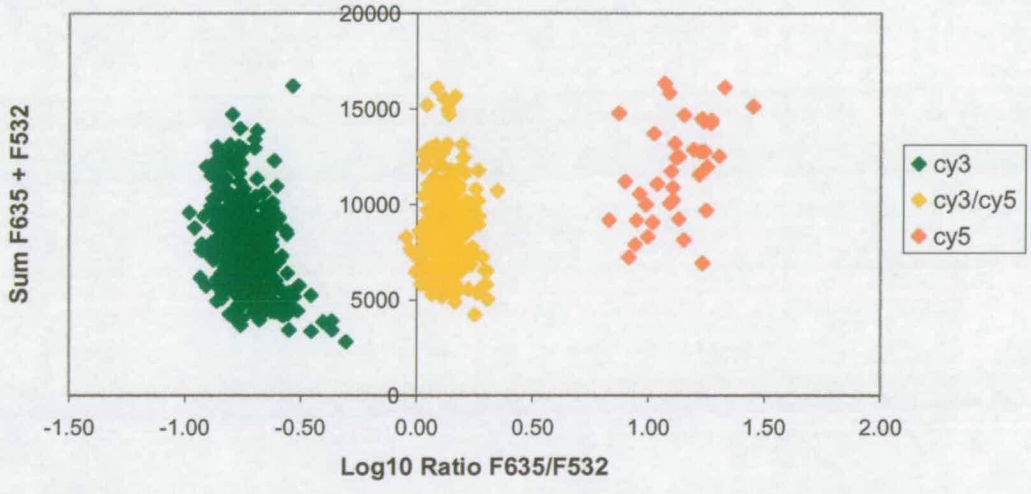


Figure 6-12 Cluster analysis of the $\log_{10}$ ratio against the sum of the mean intensity of cy5 (F635) and cy3 (F532) signals for each DNA sample, used to determine the genotype of each individual (◆).

6.3.3. Allele association analysis for SNP7.2

6.3.3.1. One-way ANOVA

The number of piglets born alive (litter size) were recorded on sows for up to five parities. A one-way analysis of variance was carried out to investigate whether there was a significant difference between the mean number of piglets born alive for sows of each of the three genotypes (*CC, CT* and *TT*) at SNP 7.2 across the five parities (Table 6-5). The genotype frequencies are also shown in Table 6-5. Graphs of the mean litter size and the 95 % confidence interval for the three genotype classes, within each parity, are shown in Figure 6-13. For animals having their second and third parity only, the number of piglets born alive for sows of genotype *TT* was significantly greater than sows of genotype *CC* ($P < 0.05$), with the heterozygous sows farrowing an intermediate number of piglets. Also as would be predicted, the number of live piglets farrowed improved as the number of parities the sow had increased (up to parity three).

Table 6-5 Mean number of piglets born alive over five parities, for sows of each genotype class at SNP7.2 (*CC*, *CT* and *TT*) and the frequency of sows of each genotype class. (n = number of sows with litter size records at each parity level).

	Mean number of piglets born alive for sows in each genotype class (frequency of sows of each genotype)			Pooled SD	P value (2 d.f.)
	CC	CT	TT		
Parity one (n = 2940)	9.79 (0.61)	10.01 (0.31)	10.25 (0.08)	3.22	$P = 0.06$
Parity two (n = 2287)	10.30[a] (0.61)	10.43 (0.31)	11.04[b] (0.08)	3.33	$P = 0.02$
Parity three (n = 1785)	10.86[a] (0.60)	10.97 (0.32)	11.77[b] (0.08)	3.45	$P = 0.01$
Parity four (n = 1051)	10.82 (0.62)	10.95 (0.30)	11.45 (0.08)	3.57	$P = 0.27$
Parity five (n = 427)	10.78 (0.69)	11.36 (0.25)	10.67 (0.06)	3.48	$P = 0.31$

[ab] Values within the same row differ significantly ($P < 0.05$)

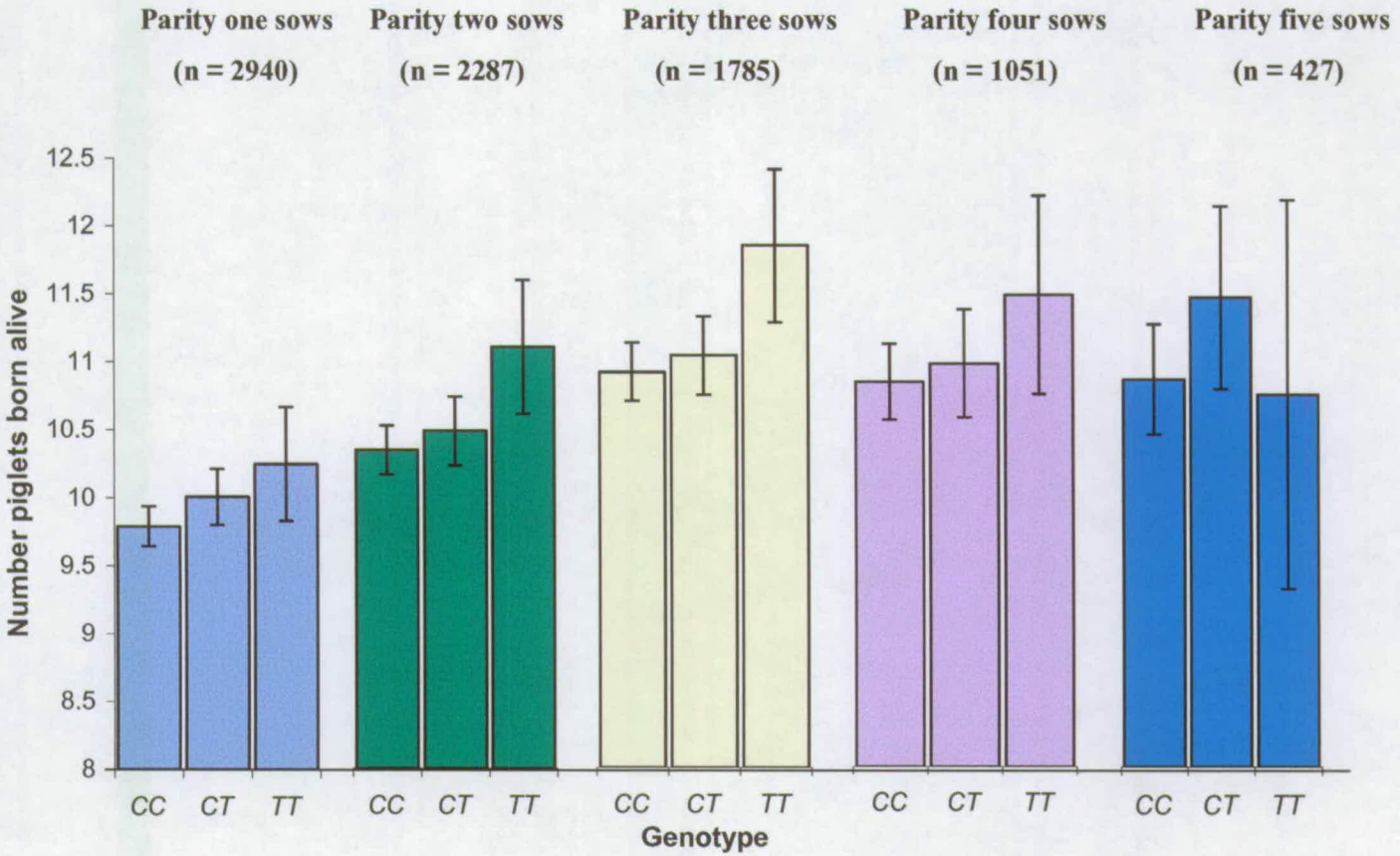


Figure 6-13 Mean number of piglets born alive for animals of genotype *CC, CT* and *TT* at SNP7.2 (encodes either a proline (base C) or a serine (base T)). Data are shown for the same group of sows over five parity records, with progressively less animals (n) in each parity group. Error bars represent 95 % confidence interval of mean (SEMx1.96) based on pooled SD.

At first inspection it appears as if there is an association between genotype and litter size for sows having their second and third parity, where sows with genotype *TT* at SNP7.2 farrow almost one piglet more than sows of genotype *CC*. All sows of purebred Meishan origin were homozygous for base *T* at this SNP and indeed this breed is one of the more prolific pig breeds. However the data set for this association analysis includes sows of varying breed origin and although the ANOVA appears to suggest that animals of genotype *TT* have higher litter sizes than sows of genotype *CC*, the analysis does not take into account other contributory factors to the phenotype. These include breed origin, the genetic background of the sow and external environmental factors such as differences in farm management.

Therefore, the mean litter size for each class of genotype within the six groups of sows of different line origin was investigated. Figure 6-14 shows the percentage of parity one sows of each of the three genotype classes at SNP7.2 (*CC, CT* and *TT*) for the six line groups of varying breed origin. Table 6-6 presents the allele frequencies for each line and the results of a Chi-square test for Hardy-Weinberg equilibrium.

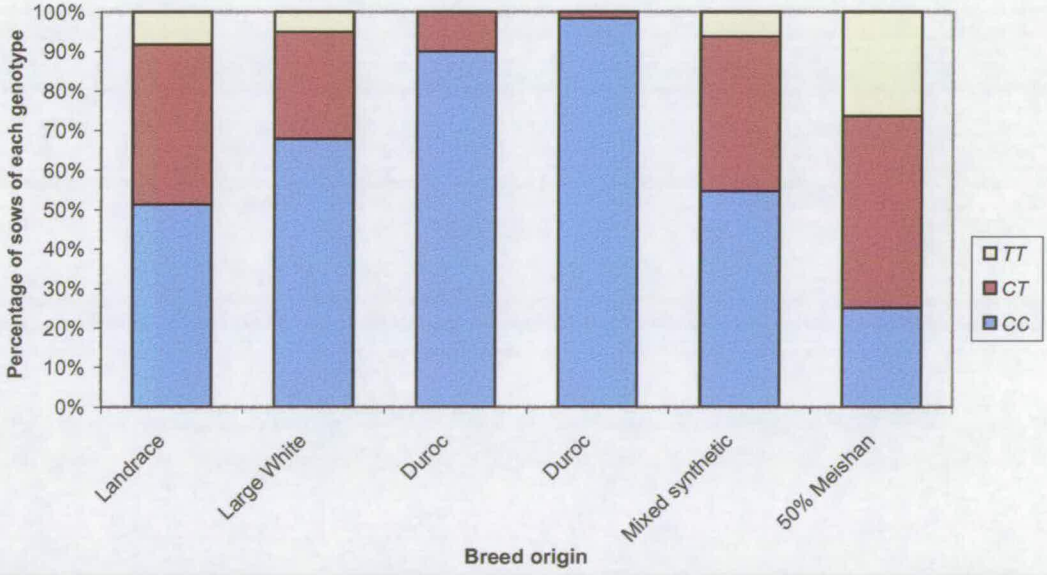


Figure 6-14 Percentage of parity one sows within each of the three genotype classes at SNP7.2 (*CC, CT* and *TT*) for the six line groups of varying breed origin.

Table 6-6 Frequencies of the alternate alleles at SNP7.2 for each line and the results of a Chi-square test for Hardy-Weinberg equilibrium

	Allele Frequencies		Test for Hardy-Weinberg equilibrium (P value)
Line	Base C	Base T	
C-Landrace	0.69	0.31	No statistical deviation from HWE (P >0.05)
D-Large White	0.82	0.18	Evidence for disequilibrium (P <0.05)
E-Duroc	0.95	0.05	No statistical deviation from HWE (P >0.05)
F-Duroc	0.99	0.01	No statistical deviation from HWE (P >0.05)
G-Mixed Synthetic	0.72	0.28	No statistical deviation from HWE (P >0.05)
H-50% Meishan	0.51	0.49	No statistical deviation from HWE (P >0.05)

None of the Duroc sows had the genotype *TT* and the largest number of sows within a line of genotype *TT*, was for the group of 50 % Meishan origin. Table 6-7 shows the mean litter size within lines for sows of each genotype class at SNP7.2 having their second and third parity (across lines, animals of genotype *CC* were shown to have significantly lower litter sizes than sows of genotype *TT* across these two parities (Table 6-5)).

Table 6-7 Mean litter size for sows of varying line origins for each of the three genotype classes at SNP7.2

	Genotype of parity two sows			Genotype of parity three sows		
Line	*CC*	*CT*	*TT*	*CC*	*CT*	*TT*
C- Landrace	10.59	10.44	10.89	11.12	10.97	11.58
D- Large White	11.03	10.02	9.97	11.44	10.16	10.90
E- Duroc	9.40	8.86	-	10.18	10.80	-
F- Duroc	9.28	6.67	-	9.55	7.50	-
G- Mixed synthetic	9.24	9.77	9.06	8.91	10.07	8.42
H- 50% Meishan	11.12	12.13	12.32	13.77	12.99	13.33

The overall trend of sows of genotype *TT* having litter sizes greater than sows of genotype *CC* is not present within each of the line groups. The only visible trend in the data in Table 6-7 is that those lines that have the highest percentage of sows of genotype *TT* (lines H and C) have larger litters. Therefore if these breeds, which are more prolific than the other breeds, have a higher number of sows of genotype *TT*, the results of the ANOVA could be explaining a breed effect on litter size. It is

certainly not the case that all sows of genotype *TT*, whatever their breed origin, have higher litter sizes than sows of genotype *CC*.

6.3.3.2. *REML analysis*

GenStat® for Windows, sixth edition was used to perform the REML (residual maximum likelihood estimation) analysis. The data set was unbalanced due to the fact that there were no sows in breed groups E or F (Duroc) of genotype *TT* at the SNP of interest. GenStat is able to take this factor into consideration within the analysis. The REML variance components analysis was used to not only test for an association between sow genotype and litter size, but also to account for the major factors affecting litter size. These included the fixed effects of farm background, breed origin, number of mating services of the sow and genotype. The identity of the sire of each of the sows was also included as a random effect. In addition, the interaction between genotype and breed origin was included in the model. The analysis was repeated for sows at each parity level.

In summary the regression model used was:

Y = farm + breed + number of services + genotype + sire of sow

Where Y = number of piglets born alive

Within REML a Wald test is used to test the statistical significance of each coefficient of the predictor variables in the model (e.g. genotype, breed or farm) on the number of piglets born alive. One of the assumptions of the analysis is that the data i.e. the number of piglets born alive are normally distributed. Within the REML analysis, it is possible to look at the distribution of the residuals and see if they follow a normal distribution. Across all the parities the distribution of the residuals was indeed very close to a straight line.

Within the Wald test a Z statistic is calculated from the coefficient of each predictor variable. This Z value is then squared, yielding a Wald statistic with a chi-square distribution and the subsequent P value is then calculated. It is only appropriate to use a chi-square distribution with large data sets, such as is the case with this experiment. With small data sets the probabilities would be underestimated. The results of the Wald tests for sows at each parity level (1-5) are shown in Table 6-8.

Table 6-8 Statistical significance of each of the predictor variables in the REML model on the number of piglets born alive at each parity level. (n = number of sows with litter size and genotype records at the SNP7.2 within *SPP1).*

Parity one sows (n=2940)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	3.23	2	$P>0.1$ (N.S.)
Line	116.01	5	$P<0.001$
Farm	0.03	1	$P>0.1$ (N.S.)
Number of services	2.62	5	$P>0.1$ (N.S.)
Genotype x Line	15.74	8	$P<0.05$

Parity two sows (n=2287)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	0.69	2	$P>0.1$ (N.S.)
Line	64.89	5	$P<0.001$
Farm	16.31	1	$P<0.001$
Number of services	5.58	4	$P>0.1$ (N.S.)
Genotype x Line	10.82	8	$P>0.1$ (N.S.)

Parity three sows (n=1785)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	1.32	2	$P>0.1$ (N.S.)
Line	68.17	5	$P<0.001$
Farm	17.39	1	$P<0.001$
Number of services	1.07	4	$P>0.1$ (N.S.)
Genotype x Line	15.76	8	$P<0.05$

Parity four sows (n=1051)

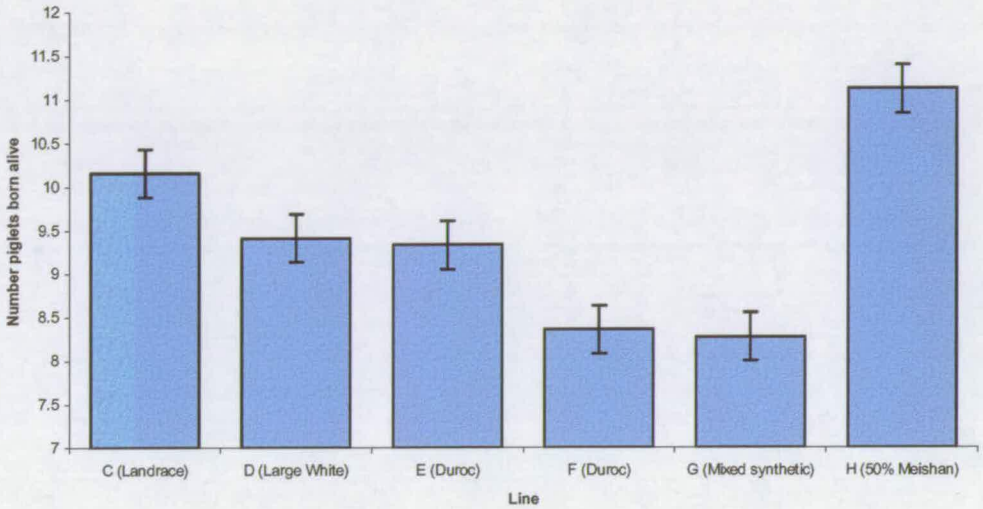
Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	1.75	2	$P>0.1$ (N.S.)
Line	26.47	5	$P<0.001$
Farm	53.14	1	$P<0.001$
Number of services	1.72	2	$P>0.1$ (N.S.)
Genotype x Line	3.70	8	$P>0.1$ (N.S.)

Parity five sows (n=427)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	1.39	2	$P>0.1$ (N.S.)
Line	32.19	5	$P<0.001$
Farm	7.69	1	$P<0.01$
Number of services	2.73	1	$P<0.1$ (N.S.)
Genotype x Line	2.33	8	$P>0.1$ (N.S.)

For parity one sows, the line or breed of the sow and the interaction between genotype and sow both had a significant effect on the number of piglets born alive. To clarify this graphs of the mean number of piglets born alive within each line and within each group of sows of each of the three genotype classes, alongside the average standard error of the differences in the means are shown in Figure 6-15.

A



B

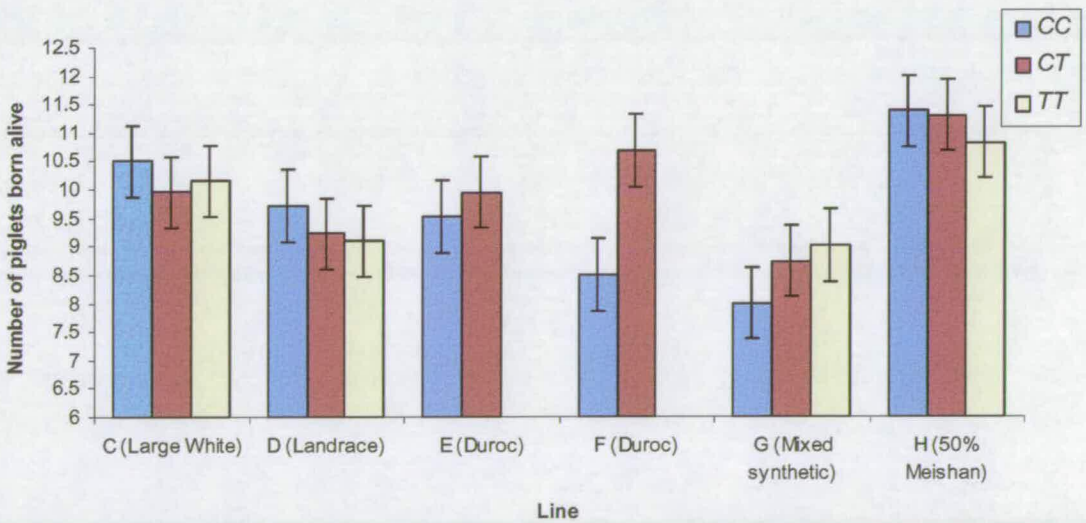


Figure 6-15 Mean number of piglets born within each line (**A**) and within each group of sows of each of the three genotype classes at SNP7.2 (**B**). Y error bars show the average standard error of the differences in the means.

The most prolific breed groups were C (Large White) and H (50 % Meishan). There is a clear significant difference in litter size between the different breed groups. Within breed groups, except group F, there are no significant differences in litter size between sows of the three genotype classes. The interaction explains that the effect of genotype varies within each of the breed groups. Across the other parities breed and farm origin had a significant effect on litter size and the interaction between genotype and breed was significant for parity three sows. When genotype had been separated out from the other influential factors, there was clearly no affect on litter size.

6.3.4. Allele association analysis for promoter SNP

Using the same PIC data set of 4017 sows and microarray genotyping method as described in section 6.2.2, the SNP at the 5' end of intron 1 (position 2770 appendix II), which resulted in a base change from guanine to thymine (described in section 5.3.2.3) was genotyped (the genotyping of this SNP was carried out by A. Day from Sygen). Genotypes were successfully ascertained at the SNP in intron 1 for 3484 sows, with phenotype records. Figure 5-14 demonstrates that this SNP site (position +110) does not lie within a region of high promoter activity or known gene regulatory regions.

If an association is found, then although this locus itself is unlikely to be a causal mutation, it may well be in linkage disequilibrium with a causual variant. Unlike SNP7.2, the SNP in intron 1 is not fixed for alternative alleles in the Large White and Meishan breeds. All the purebred Large White individuals were *TT* homozygotes and all except two of the Meishan individuals were homozygous for the base *G* allele, with these two being heterozygotes (Figure 5-7). The results of the simple one-way ANOVA did not show such a clear difference in litter size between the three genotype classes as it had done with the analysis of SNP7.2 (see Table 6-9). The proportion of sows of the three genotype classes at this SNP in intron 1 for the six breed groups was similar to that of SNP7.2 (compare Figures 6-14 and 6-16). The main difference was that six sows of the less common genotype (*GG*) at the SNP in intron 1 were present in the two Duroc lines, whereas no sows had the rare

genotype (*TT*) at SNP7.2. Table 6-10 shows the allele frequencies and results of a Chi square test for Hardy-Weinberg equilibrium for the promoter SNP.

Table 6-9 Mean number of piglets born alive over five parities, for sows of each genotype class at promoter SNP (*TT*, *TG* and *GG*). Also shown in the frequency of sows of each genotype class. (n = number of sows with litter size records at each parity level).

	Mean number of piglets born alive for sows in each genotype class (frequency of sows of each genotype)			Pooled SD	P value (2 d.f.)
	TT	TG	GG		
Parity one (n = 3484)	9.76[a] (0.65)	10.02[b] (0.29)	10.15 (0.06)	3.24	$P = 0.04$
Parity two (n = 2706)	10.34 (0.65)	10.37 (0.29)	10.77 (0.06)	3.35	$P = 0.31$
Parity three (n = 2129)	10.90 (0.66)	10.94 (0.28)	11.09 (0.06)	3.43	$P = 0.84$
Parity four (n = 1268)	11.02 (0.70)	10.98 (0.25)	11.18 (0.05)	3.53	$P = 0.91$
Parity five (n = 503)	10.75 (0.77)	11.42 (0.19)	10.00 (0.04)	3.41	$P = 0.11$

[ab] Values within the same row differ significantly ($P < 0.05$)

Table 6-10 Frequencies of the alternate alleles at promoter SNP for each line and the results of a Chi-square test for Hardy-Weinberg equilibrium

	Allele Frequencies		Test for Hardy-Weinberg equilibrium (P value)
Line	Base T	Base G	
C-Landrace	0.80	0.20	No statistical deviation from HWE ($P > 0.05$)
D-Large White	0.80	0.20	Evidence for disequilibrium ($P < 0.05$)
E-Duroc	0.93	0.07	No statistical deviation from HWE ($P > 0.05$)
F-Duroc	0.96	0.04	Evidence for disequilibrium ($P < 0.05$)
G-Mixed Synthetic	0.73	0.27	No statistical deviation from HWE ($P > 0.05$)
H-50% Meishan	0.61	0.39	No statistical deviation from HWE ($P > 0.05$)

A REML variance components analysis was then carried out as described in section 6.3.3.2., to investigate whether factors other than genotype at the intron 1 SNP affected litter size. The results are shown in Table 6-11.

Table 6-11 Statistical significance of each of the predictor variables in the REML model on the number of piglets born alive at each parity level. (n = number of sows with litter size and genotype records at the SNP in intron 1 of *SPP1)*.

Parity one sows (n=3484)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	0.02	2	$P>0.1$ (N.S.)
Line	116.70	5	$P<0.001$
Farm	0.19	1	$P>0.1$ (N.S.)
Number of services	4.77	5	$P>0.1$ (N.S.)
Genotype x Line	13.69	10	$P>0.1$ (N.S.)

Parity two sows (n=2706)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	0.68	2	$P>0.1$ (N.S.)
Line	96.42	5	$P<0.001$
Farm	16.31	1	$P<0.001$
Number of services	6.62	4	$P>0.1$ (N.S.)
Genotype x Line	15.59	10	$P>0.1$ (N.S.)

Parity three sows (n=2129)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	0.16	2	$P>0.1$ (N.S.)
Line	91.40	5	$P<0.001$
Farm	16.54	1	$P<0.001$
Number of services	0.64	4	$P>0.1$ (N.S.)
Genotype x Line	19.12	10	$P<0.05$

Parity four sows (n=1268)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	1.34	2	$P>0.1$ (N.S.)
Line	44.01	5	$P<0.001$
Farm	42.05	1	$P<0.001$
Number of services	1.22	2	$P>0.1$ (N.S.)
Genotype x Line	1.96	8	$P>0.1$ (N.S.)

Parity five sows (n=503)

Predictor variable	**Wald Statistic**	**d.f.**	**Chi-squared probability (P)**
Genotype	6.64	2	$P<0.05$
Line	54.39	5	$P<0.001$
Farm	4.49	1	$P<0.05$
Number of services	7.27	2	$P<0.05$
Genotype x Line	2.36	7	$P>0.1$ (N.S.)

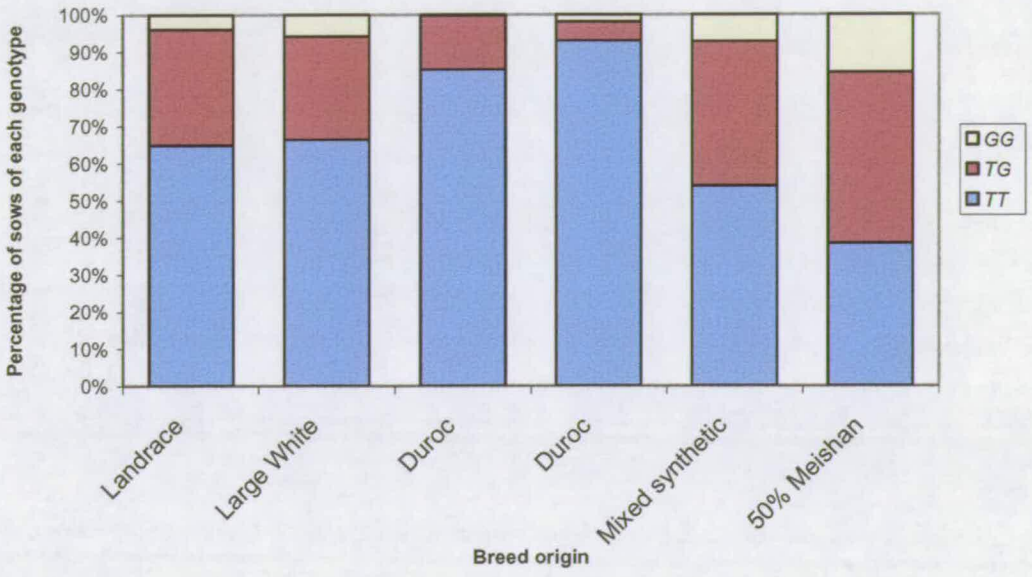


Figure 6-16 Percentage of parity one sows within each of the three genotype classes at the SNP in intron 1 (*GG*, *TG* and *TT*) for the six line groups of varying breed origin.

6.3.5. Linkage disequilibrium analysis

The genotypes at the SNP in intron 1 and the SNP in exon 7 of *SPP1* were used to test whether these two loci were in linkage disequilibrium (LD) within the six breed groups (Table 6-12). An expectation-maximisation (EM) algorithm was used to estimate the maximum likelihood of haplotype frequencies under the assumption of Hardy-Weinberg equilibrium, as described by Excoffier and Slatkin (1995). This allows the gametic phase of haplotypes for diploid individuals heterozygous at both loci, to be determined. The measured unit of linkage disequilibrium is "D", that is the deviation of the haplotype frequency from that expected at linkage equilibrium (Gavrilets, 2001). Where D'= 0, the loci are in complete gametic equilibrium and where D'= 1, the loci are in complete gametic disequilibrium. The analysis assumes a random mating structure in the population. The program used for the analysis was kindly provided by Albert Tenesa from University of Edinburgh.

Table 6-12 Evidence for gametic disequilibrium between the SNPs typed in intron 1 and exon 7 of *SPP1* across PIC sows of varying breed origin.

Line group	No. sows with genotype data at both SNPs	Significance of evidence for linkage disequilibrium	Significance of evidence of non-random mating	D'
C-Landrace	811	$P<0.0001$	$P<0.05$	0.86
D-Large White	807	$P<0.0001$	$P<0.0001$	0.95
E-Duroc	226	$P<0.0001$	$P>0.1$ (N.S.)	0.96
F-Duroc	221	$P>0.1$ (N.S.)	$P<0.01$	0.28
G-Mixed synthetic	284	$P<0.0001$	$P>0.1$ (N.S.)	0.98
H-50 % Meishan	301	$P<0.0001$	$P>0.1$ (N.S.)	0.87
TOTAL	**2650**	***P*<0.0001**	***P*<0.0001**	**0.89**

For all the breed groups, except F, there is strong evidence that these two SNPs, which are around 6 kb apart in the *SPP1* gene, are in linkage disequilibrium. If other loci within *SPP1* or nearby are also found to be in LD with these two SNPs, then it can be confidently predicted that they will also show no association with litter size.

It was only possible to test for linkage disequilibrium, by assuming that the populations are under random mating and it can be seen from Table 6-12 that lines C, D and F showed evidence of non-random mating. Because of the difficulty of determining haplotypes for heterozygous individuals at both loci, a model allowing for non-random mating will always produce a perfect fit to the observed genotype counts and therefore the analysis leaves no degree of freedom to test for linkage disequilibrium (Sham, 1998). Therefore those analyses that show evidence for non-random mating have to be interpreted with caution.

6.4. Discussion

6.4.1. Confirmation of SNPs

The most common form of DNA sequence variation is single nucleotide polymorphisms (SNPs). An operational definition for a polymorphism (including SNPs) is genetic variation for which the rare allele/variant has a frequency of at least 1 % in the population of interest (Gut, 2001). SNPs occur on average every 1000 bp along the 3 billion nucleotides of the human genome (Marth et al., 1999). Sachidanandam *et al.* (2001) identified 1.42 million SNPs and as expected they found much less variation in coding regions than non-coding regions. They found 60,000 SNPs in exonic gene regions, which resulted in non-synonymous amino acid changes, at a frequency of 1 SNP per 1.08 kb of coding sequence.

Out of the eight sequence variants of interest investigated in this chapter, all of them except SNP7.1 (Figure 6-1) were confirmed in genomic DNA. Although closely linked SNPs will often be found in linkage disequilibrium (LD), Kruglyak (1999) predicted that haplotype blocks would be limited to around 3 kb in humans and that at least 500,000 SNPs would need to be typed for a whole genome association study. There are currently around 2.4 million unique human SNPs in the public databases and several studies are involved with mapping haplotype blocks within the human genome to aid the search for causal SNPs involved with complex traits (Couzin, 2002). Similar studies in the mouse are also now underway (Lindblad-Toh et al., 2000), in particular following the recent completion of the draft sequence of the mouse genome (Waterston et al., 2002). Where a SNP contributes to a complex trait, all the SNPs within the haplotype block will also show association with the trait. Therefore only one SNP needs to be typed within each haplotype block and where that block is as small as a few kb, a non-causal SNP within the same gene as the causal allele, can often also be associated to the trait (Dawson et al., 2002). In the same way, it was hoped with this study that if an allele at the SNP typed over the large commercial population did show an association with litter size, that it would either be the causal SNP itself or in LD to a nearby causal SNP within the same gene.

There was seen to be strong evidence that the SNPs in intron 1 and exon 7 typed over the group of PIC sows are in linkage disequilibrium. Therefore any other nearby loci also found to be in LD with these two SNPs are also unlikely to show an association with variation in litter size in these sows.

Recently it is becoming more common to utilise association studies of haplotype blocks rather than classical linkage mapping to locate genes/loci controlling complex traits. Linkage mapping relies on recombination events within structured families and therefore the regions of LD tend to be relatively large. Whereas with association based analyses the haplotype blocks can be as small as a few kb (Cheung and Spielman, 2002). SNP markers are much more stable within a population than polymorphic microsatellite markers (Sachidanandam et al., 2001). They are therefore almost exclusively Identical By Descent (IBD) and can even be used for QTL scans across pedigrees of distantly related animals (Fahrenkrug et al., 2002). These SNP haplotype maps are certainly more common in human studies where structured pedigree crosses are not possible. However Fahrenkrug *et al.* (2002) and McEwan *et al.* (2002) also describe the emerging importance of SNP maps for fine mapping of QTL regions in livestock species. High-density SNP marker maps for fine mapping require a much greater number of markers to be typed and as a consequence there is a requirement for high throughput genotyping.

6.4.2. Use of microarray technology for genotyping SNPs

The technology for mutation detection has recently progressed rapidly alongside the advances in genome research, in particular whole genome sequencing projects (Vaughan and McCarthy, 1998). Various high throughput genotyping methods are now available for scoring bialleic loci, including several that use microarray technology. When handling such large datasets the main requirements of the system of analysis are that they are accurate, rapid and cost effective (Fan et al., 2000). Examples of common genotyping methods of SNPs not involving microarray include the TaqMan® fluorescence assay, primer extension, oligonucleotide ligation or RFLP (for more detail on these assays see Sayers *et al.* (2000) and the review by Gut (2001). It is also possible to use comparative sequencing, where the bi-alleic

sequences are compared for different individuals. Where an individual is heterozygous for a SNP, a multiple peak for the two alternative nucleotides will be seen (Fahrenkrug et al., 2002).

The type of genotyping method chosen depends on whether the study requires the genotyping of a limited number of SNP markers in a large population or the analysis of a large number of SNP markers in one or a few individuals (Gut, 2001). For the typing of a few SNPs over hundreds or even thousands of individuals the use of microarray chips is ideal. Iwasaki *et al.* (2002) found the success rate of genotyping SNPs using their microarray based allele specific oligonucleotide (ASO) hybridisation method to be very high (99.9 %). Because this group were testing several SNPs over a small number of individuals they arrayed the allele specific probes of 11 bp in length to the glass slide and hybridised with PCR products of around 60 bp in length from the DNA samples.

This methodology is reversed with respect to the genotyping method I used, and also as described by Ewen *et al.* (2002). For typing a few SNPs over a large number of animals, the PCR amplicons were arrayed onto the slide and hybridised with short labelled oligonucleotide probes. Iwasaki *et al.* (2002) also used a relatively low hybridisation temperature of around 30° C (34° C was used for my study) and of most interest was their incorporation of oligonucleotide "chaperons" into their experimental design. This is where two separate oligonucleotides are designed to hybridise to both ends of the 60 bp PCR amplicon, leaving only the 11 bp region specific to the allele-specific probes on the glass slide exposed. This is a very effective way of ensuring that the chemiluminescence signals produced refer to the alternative alleles of the SNP of interest and not to another nucleotide site within the PCR amplicon. One of the main problems of hybridisation is the chance of non-specific binding and it might well be worth incorporating "molecular chaperons" to the experimental design used here to type SNPs within *SPP1*.

The allele specific probes used in this study were all around 15 bp in length, with the SNP site designed to be in the centre of the probe (see Table 6-4), in order to

increase the chance of specific binding. There is still the possibility however that because the probes for the alternative alleles were so similar in sequence i.e. only one base pair difference, that the binding to the PCR amplicons on the slide may not always be specific. However the stringent wash conditions used here should have ensured that any hybridisation probe that bound to DNA that was not fully complementary was removed (Gut, 2001). In addition every DNA sample was spotted onto the slide in triplicate and the intensity of fluorescence for the two alternative fluorophores checked to be similar for all three individual array spots and then the mean value calculated across them.

Every slide contained a negative control sample in order to calculate the background fluoresence of the slide, however I feel that it would have been useful to have also incorporated a positive control sample on every slide. An obvious choice for this would be to use DNA from the F1 Meishan x Large White boar used to create the BAC library. This DNA had been used to detect the alternative alleles at the SNPs within *SPP1* and therefore both the cy3 and cy5 fluorophores should fluoresce with an equal intensity and confirm that the individual was heterozygous. It may also be worthwhile including DNA from two individuals, each known to be homozygous for the alternative alleles. Therefore the intensity of the signal for all the DNA samples from sows of unknown genotype could have been compared to the three controls.

It is likely that the main reason that genotypes could not be determined for over 1000 DNA samples was due to the high levels of background noise on the slides (see section 6.3.2). The conditions outlined for the classification of genotypes had been strict to ensure that those typings that were considered successful were included (section 6.3.2). However it is far from ideal to lose over 25 % of the available data. It was initially thought that perhaps the DNA concentration of the 4017 samples had not been sufficient. However the exact concentrations were not known and it is possible that they may well have varied across samples, which could explain why some spots fluoresced more brightly than others. Other suggestions to reduce the high background noise include baking the slides for longer than two hours and possibly using UV cross-linking in order to help the DNA to bind more efficiently to

the slides. However the protocols derived from those of the Cold Spring Harbor laboratory (found at http://www.microarrays.org/pdfs/PostProcessing2001.pdf) suggest that although UV cross-linking will enhance binding of long oligonucleotide DNA, it has no effect on the binding of short PCR products to the glass slide.

There may well have been batch variation in the poly-lysine slides used; this would explain why the first set of slides worked better than the later ones. One could use a higher quality of slide, which has a surface coating with very low intrinsic fluorescence and therefore maximises the signal to noise ratio. These are more expensive but may be worthwhile as only a few slides were required to be printed for this experiment. Furthermore the Erie Scientific Company demonstrated that the use of DMSO as a printing buffer results in good spot morphology and high signal intensity http://www.eriesci.com/tech_info/DNA_proto.html. They also describe that an additional advantage of DMSO is that the DNA spots do not dry out so easily, a problem, which causes do-nut shaped spots (as described in the testing of hypothesis ii) in appendix III). Also the array post-processing protocol (found at http://www.microarrays.org/pdfs/PostProcessing2001.pdf) describe the use of re-hydration of the DNA spots to eliminate the do-nut shape and to increase the amount of total DNA bound to the slide after processing. These are all possible ways of improving the intensity of fluorescence resulting from hybridisation with the allele specific probes.

Many protocols for binding DNA to poly-lysine slides describe the need to use a succinic anhydride block to cap exposed amines on the slide, to prevent binding of probes to non-specific sites on the slide (for example see http://www.microarrays.org/pdfs/PostProcessing2001.pdf). A BSA block was tested within this study (Appendix III). Unfortunately it reduced the overall spot intensity and consequently increased comparative background noise. It was therefore decided that a block would not be used.

Finally, it would have been useful to have confirmed, by gel separation, that the PCR reactions across the 4017 DNA samples had been successful. One possible way to

do this is to use microplate-assay diagonal-gel electrophoresis (MADGE) of the PCR products contained within 96 well plates (Sayers et al., 2000). This would have involved transferring the samples from the 384 well plates to 96 well format, and if possible, it would be better to run all 384 samples together. In addition the DNA should ideally have been purified from the PCR products using for example filtration plates in order to remove the PCR primers and increase the specificity of the target DNA on the glass slide. Both these steps would however be time consuming and very costly for such a large number of DNA samples.

6.4.3. Association studies in large porcine commercial breeding populations

The number of successful genotypes obtained at SNP7.2 was sufficient to carry out an allele association analysis within the PIC commercial breeding populations. With the simple one-way ANOVA it appeared as if there was an allele substitution effect with SNP7.2 and litter size for second and third parity sows. Animals of genotype *TT* appeared to farrow almost one piglet more than sows of genotype *CC* (Figure 6-13). Interestingly the frequency of the allele which encodes proline within the Large White, Landrace and Duroc breeds was between 0.72 and 0.99 and within the sixth breed line which was of 50 % Meishan origin the frequency of the allele encoding a proline was only 0.49. However, once the external factors such as breed origin and the effects of environment were taken into consideration it was realised that the difference in litter size observed between sows was in fact explained by the breed origin. In some cases differences were also explained by external environmental influences on the phenotype and there was in fact no association found between the genotype of the sow at SNP7.2 and litter size. It can be concluded that there are other loci at unknown locations in the genome that control the differences in litter size between various pig breeds. This clearly demonstrates that with association analyses it is important to incorporate all possible influential factors on the phenotype into the model used, to ensure that the effect seen is genuine and the results observed are not confounded with other variables.

The association analysis of genotype at the SNP in intron 1 and litter size also revealed no significant association. The REML analysis again concluded that it was mainly the difference in breed origin of the sow lines, which explained the variation in litter size. It can be seen from Tables 6-8 and 6-11 that the variables having a significant effect on the number of piglets born alive are almost identical across the first four parities; the only difference being that there is no significant interaction between genotype and phenotype in the parity one sows genotyped for the SNP in intron 1. Interestingly there appears to be a significant effect of genotype at the intron 1 SNP on the litter size records of parity five sows (Table 6-11). However, the number of sows in this group is much smaller than the others, therefore more greatly reducing the statistical power of the test, and there is no obvious biological explanation why there would be an association between genotype and phenotype only at the sow's fifth parity.

When conducting the REML analysis it was decided that the most accurate method was to carry out separate analyses at each parity level. The litter size data could have been pooled across parities to give a much larger data set and the sow fitted as a random effect into the model, to allow for each sow having more than one parity record. However, Alfonso *et al* (1997) discuss that the genetic background of litter size is different for each parity i.e. the genetic correlation between the first and subsequent parities is less than one. Additionally, there were believed to be a sufficient number of animals (up to 3484 sows per parity) to give the power required for an accurate association analysis within parity.

In the future it may be worthwhile extending the current work by conducting a REML analysis for each breed separately, at each of the parity levels. This would determine whether there was an association between specific alleles at the SNP of interest and litter size, for each of the breeds. Although with this type of analysis, the numbers of animals available for each analysis is much lower and there may therefore not be sufficient power to detect an association.

There have been several recent examples of association studies of single candidate gene loci in livestock species for traits of economic importance. An example was the demonstration of an allele substitution effect of alleles (3 out of 8) within the microsatellite locus of *SPP1* associated with litter size (microsatellite repeat maps 5' of exon 1 between position 2442 and 2505 appendix II). This study was carried out on a PIC line of 50 % Meishan origin and the genetic difference observed was an additive effect of between 0.41 and 0.95 piglets per allele copy. They used a mixed linear regression model which included effects of genetic line, herd-year-season of farrowing, parity and service type (natural or AI). van der Steen *et al.* (1997) suggested that the *SPP1* microsatellite marker must be linked to a QTL of litter size and that if *SPP1* is not the causal gene itself, then there could be a causal loci in LD with the microsatellite locus. The QTL study carried out by myself (chapter 2) was the first to identify a QTL for prenatal survival and the related trait of litter size around the *SPP1* gene.

An additional locus within *SPP1*, the SINE in intron six, has also been proposed as a marker for litter size. SINEs or short interspersed nuclear elements were originally discovered in humans (Alu repeats) (Moran, 1998). The family of SINEs with the PRE-1 (porcine repetitive element), as was seen within intron 6 of *SPP1* are the most common with an estimated 100,000 copies throughout the pig genome (Singer et al., 1987). They are derived from small cellular RNA species, which have been reversed transcribed and then integrated into the genome (Moran, 1998). Interestingly these PRE-1 SINEs are absent from bovidae, mice and humans (Takahashi et al., 1992) and are present at a similar level of abundance in warthogs and peccaries as domestic pigs (Yasue and Wada, 1996). This implies that the repeat originated after the divergence of pigs from ruminants. In effect, these repeat elements are processed pseudogenes which are non functional (Moran, 1998). Therefore the association found by Korwin-Kossakowska *et al.* (2002) between the presence of a SINE within *SPP1* in second and later parities of Polish commercial sows and litter size is likely to be explaining association of a marker in LD with the SINE and not a causal role of the SINE itself.

The frequency of the presence of the SINE within their population was 0.21. Knoll *et al.* (1999) determined the frequency to be 0.57 in Landrace, 0.26 in Large White, 0.16 in Duroc and 0.82 in Pietran breeds. It can be seen from Figure 6-2 that all of the 15 purebred Meishan animals were homozygous for the absence of the SINE and the frequency of the presence of the SINE in the 14 purebred Large White individuals was 0.54.

Once beneficial alleles at specific loci have been shown to be associated with a positive improvement in a livestock trait, it is possible to introduce marker-assisted selection (MAS) within commercial breeding populations (Georges, 1999). By integrating molecular genetics approaches into traditional selective breeding methods, it should be possible to achieve the maximum improvement in the economic value of domesticated livestock populations (Lande and Thompson, 1990). However for MAS to be successful, it is important that beneficial alleles for a given trait are shown to have no deleterious pleiotropic effects with other performance traits (Rothschild et al., 1997).

Unfortunately due to overinterpretation of results and poor study design the literature is teeming with reports of associations that either cannot be replicated or for which corroboration by linkage has been impossible to find (Cardon and Bell, 2001). However fine mapping of disease trait loci through a combination of linkage and association analyses remain important strategies for the identification of causative gene variants, particularly where the trait is influenced by one or more polymorphisms within the same gene (Soubrier et al., 2002). One of the main goals of human genetics is to understand how phenotypic variation seen in normal and clinical contexts is related to the underlying sequence variation in the genome (Cheung and Spielman, 2002). Most association studies tend to investigate single loci that have been identified as putative candidate genes. In contrast SNP-based genome-wide association analyses are possible and in theory relatively straightforward. There are still however many technical and statistical obstacles. The technique needs to be reliable and efficient to type a large number of markers over the individuals of interest. Genomic DNA is complex and repetitive and it can

therefore be difficult to analyse markers by hybrid-based array technology due to the problems of cross hybridisation (Cheung and Spielman, 2002). There are now microarrays available to genotype a few thousand SNPs over the human genome (http://www.affymetrix.com and http://www.illumina.com). Because the exact size of haplotype blocks is unknown and the size of each block varies considerably depending upon whether it is found within a recombination hotspot (Miller and Kwok, 2001), it is difficult to know the exact density of markers required for a whole genome association study. The difficulties encountered with the analysis is that it must be able to detect small contributions to the phenotype from several genes or loci, whilst allowing for the inherent false positive error rate involved when testing such a large number of markers. It is also important to consider gene-gene or gene-environment interactions on the control of complex traits (Cheung and Spielman, 2002). The problem with quantitative characteristics is that where genes or loci are identified to have a relatively small effect on that trait it can be difficult to ascertain its exact role within what can be a complex pathway of physiological control.

In summary there was no association found between alleles at the candidate causal non-synonymous mutation in exon 7 of *SPP1* and at the SNP in intron 1 and the number of piglets born alive recorded in around 3000 sows of varying breed origin. Although there is evidence from several independent studies suggesting *SPP1* as a physiological and positional candidate gene involved in the variation seen in prenatal survival levels and litter size between pig breeds, it may of course be another gene within the QTL region identified on SSC8qter that is responsible for this variation. It would be worth investigating other genes identified in chapters 3 and 4 to map within the QTL region, such as *IBSP* and *SPARCL1,* for a physiological role.

Chapter Seven

7. CONCLUSIONS AND FUTURE WORK

Previous studies by various groups have highlighted regions on SSC8 associated with variation in ovulation rate (Rohrer *et al.*, 1999; Rathje *et al.*, 1997; Wilkie *et al.*, 1999; Braunschweig *et al.*, 2001 and Jiang *et al.*, 2001, teat number (Cassady et al., 2001), age at puberty (Cassady et al., 2001) and litter size (van der Steen *et al.*, 1997 and (Korwin-Kossakowska et al., 2002) in various pig breeds. Therefore this project focussed on a detailed investigation of porcine chromosome 8 (SSC8) in particular, to search for an association between loci on this chromosome and the variation seen in female reproductive performance between a European commercial breed, the Large White and the highly prolific Chinese Meishan breed.

Prior to the 1990s there was seen to be a very low selection response in sow fertility traits by selective breeding alone. However, during the last 12 years or so there has been a significant increase of an average of 0.1 to 0.2 piglets per year, as a result of the application of Best Linear Unbiased Prediction (BLUP) and the use of breeding information from families of large populations (Merks et al., 2000). However, if loci can be identified to be associated with improved prolificacy or ideally if the underlying causal genetic mutation can be discovered, then marker assisted selection or introgression can be used more successfully to increase the frequency of desirable alleles. Care must be taken to ensure that the beneficial alleles are not in linkage disequilibrium with undesirable alleles for other traits. In this way sow fertility can be improved directly using a targeted technology that is much more successful than selective breeding relying on information from just the animal's phenotype. This strategy is particularly important for traits of low heritability; for example the heritability of litter size in pigs is typically between 0.05 and 0.15 (Alfonso et al., 1997).

The QTL scan of SSC8 revealed a clear QTL region associated with prenatal survival around the telomere of the q arm, in a group of young sows, which was significant at the equivalent of the genome-wide level (P <0.05). This was co-located with a QTL for the related trait of litter size, which was statistically significant at the nominal

level (P <0.01). In addition the analyses revealed a QTL for teat number around the centromere of the chromosome, also significant at the equivalent of a genome-wide level (P <0.05). The additive genetic effect for all three of these QTL indicated increased performance from the Meishan breed.

There was evidence to suggest that there were in fact two QTL for teat number on SSC8, acting in opposite directions. Indeed when the genotypes at the *AREG* locus were fitted as fixed effects into the QTL analysis this removed the effect of the QTL near *AREG* and improved the statistical support for the second QTL around the *SLIT2* gene (see Figure 2-10). With the extended QTL models, there was also seen to be limited support for a QTL for ovulation rate around the centromere, near the *GNRHR* gene, and also near the telomere of the p arm (see Figure 2-8). This result is supported by the findings of (Rohrer *et al.*, 1999; Wilkie *et al.*, 1999; Braunschweig *et al.*, 2001 and (Jiang et al., 2001).

When this project was initiated the number of genes mapped to SSC8, in particular around the telomere of the q arm, was limited. Therefore information from a comparative gene map of human chromosome 4 was used to construct a detailed map of SSC8. These two chromosomes are highly conserved across species. The technique of radiation hybrid mapping was used to order the genes on the chromosome relative to a framework of markers already known to map to SSC8. Additional markers and genes of interest can be rapidly added in the future to this detailed framework map.

When searching for physiological candidate genes in the QTL regions of interest, the radiation hybrid map of SSC8 served as a useful tool. Not only were ten novel porcine homologues of human genes mapped to SSC8, in addition the map can now be confidently aligned with the sequence and gene map of HSA4 and many human genes predicted to map to specific regions on SSC8. For example the gene order at the telomere of the q arm was seen to be inverted relative to HSA4 and Figure 4-7 demonstrates how one can confidently predict that the twenty-three human genes on the equivalent region on HSA4 will map within the prenatal survival QTL region.

I would have liked to have mapped the genes *HNRPD* and *BMP3* in the pig, in order to refine the exact location of the breakpoint of the region that has been inverted; however it had not been possible to optimise the PCR conditions for primers designed within these genes, to map them to the radiation hybrid map.

During this study I focussed on the comparison between human and pig only. The main reasons for this were because HSA4 and SSC8 are so highly conserved and also the number of genes mapped in humans is large and the genomic sequencing effort was well advanced. As Figure 4-5 shows, the area of the prenatal survival QTL could be predicted to map to the q arm telomere of mouse chromosome 5 and maybe parts to mouse chromosome 3 and also to sheep and cattle chromosomes 6. It would be interesting in particular to align the mouse gene map and recently completed draft genome sequence with that of the pig, to search for additional homologous candidate genes. Especially as the reproductive physiology of this species is more similar to the pig, in terms of litter size, than it is to the human. However, unlike mice and humans, implantation in the sheep, cattle and pigs is non-invasive (Johnson and Everitt, 1995).

QTL studies investigating prolificacy traits have been reported for mice and cattle. However none were seen to map to MMU5, MMU3 or BSA6. Significant QTL for ovulation rate have been located on bovine chromosomes 7, 19 and 5 (Kappes *et al.*, 2000; Kirkpatrick *et al.*, 2000; Lien *et al.*, 2000 and (Blattman et al., 1996)) and a significant QTL was detected on mouse chromosome 2, where individuals inheriting both allele copies displayed an increase in embryo survival of 13.6 % (Pomp et al., 1995). In addition Spearow *et al.* (1995) mapped QTL for ovulation rate on mouse chromosomes 2, 6 and X.

I concluded that although the gene *SLIT2* was seen to map to one of the teat number QTL on SSC8, this gene is an unlikely candidate because it is a homologue of a Drosophila gene, which plays a role in central nervous system midline formation during embryogenesis. The gene *STE* (estrogen-preferring sulfotransferase) maps within the second QTL region for teat number at around 100 cM and Jiang *et al.*

(2002b) suggested an association between a marker for this gene and teat number in the Roslin Meishan x Large White populations. However the role of STE is to remove the sulfate group from the precursor estrone sulfate to form active estrogens and (Bernier *et al.*, 1994 and (Her et al., 1995) demonstrated that *STE* expressed in the placenta and brain maps to human chromosome 16 and it is STE from the liver which maps to human chromosome 4.

MAN2B2 (alpha mannosidase 2B2) was mapped to the telomere of the p arm of SSC8 and is a strong candidate controlling the variation seen in ovulation rate at a QTL mapped to this region (Rohrer et al., 1999), however Campbell and Rohrer (2000) found no association between a polymorphism found in the putative start codon of this gene and ovulation rate in Meishan and White composite breeds of pig. *GNRHR* (gonadotrophin releasing hormone receptor) is a strong candidate for the ovulation rate QTL mapped to the centromere of SSC8 (Wilkie *et al.*, 1999 and (Braunschweig et al., 2001) and indeed Jiang *et al.* (2001) reported an association between the allele most prevalent in the Meishan breed at a polymorphism within this gene and increased ovulation rate in sows having their first parity.

The homologue of the sheep Booroola gene, located on sheep chromosome 6, was predicted to map within the prenatal survival QTL region. The gene has a major effect on litter size in sheep, of around one extra lamb per copy of the gene (Montgomery et al., 1992). However, in sheep this gene improves litter size through ovulation rate and not prenatal survival and in addition the comparative mapping study revealed that the porcine homologue of the Booroola gene in fact maps outside of the QTL region.

Out of the five genes now mapped within the prenatal survival/litter size QTL region (*DSPP, IBSP, COPS4, SPARCL1* and *SPP1*), SPARC-like 1 and secreted phosphoprotein 1 are the only known physiological candidate genes involved in the control in variation in prenatal survival between pig breeds. *SPARCL1* has a role in cell adhesion and is expressed in the ovary and placenta (Girard and Springer, 1995). This gene also maps within the age at puberty QTL mapped by Cassady *et al.* (2001)

and the Nebraska group found differential expression of the gene in the anterior pituitary of a control group of pigs and a line selected for increased ovulation rate and embryo survival over several generations (Bertani et al., 2002) and they are currently investigating this gene further.

Secreted phosphoprotein 1 (*SPP1*) has a role in controlling embryo implantation and placentation. It is known from physiological studies that the Meishan breed has a larger litter through improvements in prenatal survival levels and not through an increased ovulation rate. Indeed during the peri-implantation period (days 12-18 of gestation), the Meishan breed displays a significant reduction from the 75 % of embryo loss seen in commercial European breeds such as the Large White (Ford, 1997). During days 5-12 of gestation, the Meishan embryos secrete less oestradiol than Yorkshire embryos (Ford, 1997). The uterine epithelium is sensitive to the oestradiol and subsequently the level of uterine endometrial secretions in Meishan sows is reduced. This reduced histotroph secretion acts on the conceptuses in the uterus and results in a decreased litter variation in embryo length, weight and distance between attachment sites, therefore allowing a greater proportion of the litter to survive the gestation period (Geisert et al., 1982). In Large White breeds, the larger more advanced embryos alter the uterine environment to the detriment of the smaller embryos (Pope et al., 1990). In addition the Meishan fetuses trigger an increased placental efficiency during late gestation, allowing the embryos a smaller area within the uterus to develop and therefore reducing competition between them (Biensen et al., 1999).

There has been shown to be an increase in *SPP1* mRNA expression from the luminal epithelium of the endometrium of pigs after day 15 of gestation (Garlow et al., 2002). The biologically active 45 kDa form of the protein is released as part of the histotroph secretion from the endometrium and binds to vitronectin receptors on the trophectoderm of the blastocyst, resulting in elongation of the conceptus and also inducing adhesion and cell signalling between the conceptus and the endometrium, vital for attachment, implantation and placentation (Johnson *et al.*, 2000 and (Johnson et al., 1999b). This control in peri-implantation loss from the sow

correlates with the results of the QTL study, where the difference in prenatal survival level was associated with the maternal genotype.

van der Steen *et al.* (1997) found an association between alleles at the microsatellite locus of *SPP1* (5' of coding region of gene) and an increase in litter size in a Meishan (50 %) synthetic line. Korwin-Kossakowska *et al.* (2002) found an association between the presence of a SINE (Short Interspersed Element or mammalian-wide interspersed repeat) in intron 6 of the gene and an increase in litter size in a commercial Polish pig line. van der Steen *et al.* (1997) state that any locus found to be associated with the trait may not necessarily be the causal mutation, but may be in linkage disequilibrium with and therefore explain the causal mutation, which may be located within *SPP1* or even within an unknown neighbouring gene.

Therefore this project focussed specifically on *SPP1* as a physiological and positional candidate gene for the control of prenatal survival variation seen between the Meishan and Large White breeds. The entire sequence of the gene, including 2 kb 5' of exon 1 was deduced from Meishan and Large White origin and putative casual variants responsible for the variation in prenatal survival identified. The comparison of the 10 kb sequence of *SPP1* from Meishan and Large White origin revealed eighty-five single nucleotide polymorphisms (SNPs), ten sequence insertions or deletions, the microsatellite tandem repeat and the presence or absence of a SINE.

All variants of interest were typed over genomic DNA from the Roslin Meishan x Large White populations to confirm that they were genuine polymorphisms and not sequencing errors or artefacts. Six SNPs were discovered in exons, three of which encoded a non-synonymous amino acid change. The two in exon 6 encoded a change from an alanine to a threonine and a valine to an alanine, both of which are relatively conserved changes. The third in exon 7 encoded a change from a proline to a serine. None of these three changes in amino acid altered the predicted secondary structure of the protein.

The presence or absence of the SINE and the amino acid change from an alanine to a threonine were typed over all individuals in the three Roslin Meishan x Large White populations and when the genotypes at these two loci were included into the QTL analysis as fixed effects, neither explained any additional variation in litter size or prenatal survival over and above the effect of the QTL itself. A marker for the SNP in exon 7 had already been included in the original QTL analysis.

SPP1 has a significant number of potential serine and threonine phosphorylation sites Safran *et al.* (1998) and Yamamoto *et al.* (1995) suggest that transcription from the SPP1 promoter is normally repressed or controlled at the posttranscriptional level. There is evidence that decreased phosphorylation of SPP1, leads to a reduced binding to cells with vitronectin ($\alpha_v\beta_3$) receptors (Safran et al., 1998). The serine, encoded by the allele at the SNP in exon 7, fixed in the Meishan pure-breds, lies within a casein kinase II phosphorylation site, which is disrupted by the proline, encoded by the alternative allele at the SNP, which is fixed in the Large White pure-bred animals. The serine amino acid residue was found to be conserved across the Meishan breed of pig, human, rabbit, mouse, rat, cattle and sheep, whereas the proline was only present in the Large White and Landrace breeds of pig. This high level of conservation most likely indicates an important function of this site within the protein.

The only variant of significance within the promoter regions was the microsatellite repeat and the function of these repeats in the porcine genome is unknown. The main sequence variant of interest was the SNP in exon 7, which encodes a change from a proline to a serine. The final experiment in this project investigated whether alleles at this SNP associated with a difference in prenatal survival and the resulting trait of litter size in the pedigree experimental crosses, were segregating within commercial breeding populations. I was fortunate to have access to a large dataset of over 4000 sows of varying breed origin with litter size records over several parities and corresponding DNA samples for genotyping. This number of animals should have given sufficient statistical power to detect any association between genotype and phenotype. Unfortunately it was not possible to test the trait of prenatal survival

directly as ovulation rate was not recorded for these sows. However litter size is a product of prenatal survival and it is therefore valid to search for an association between genotype of sows of varying breed origin and the trait of great commercial interest, litter size.

The method of microarray genotyping was used to type these bialleic SNPs over DNA from all the sows. The SNP within intron 1 was independently genotyped and a total of 2650 sows were successfully typed for both of these SNPs. Using REML variance components analysis, there was found to be no association between the genotypes at either of these loci within *SPP1* and litter size. It was found that as expected, there was a highly significant association between the breed origin of the sows and their litter size records. Interestingly the sows of 50 % Meishan origin had a much higher frequency of the less common allele at both SNP sites (base T at the SNP in exon 7 and base G at the SNP in intron 1), than the sows of Landrace, Large White, Duroc and mixed synthetic breed origin. All the purebred Meishan animals were seen to be homozygous for the allele *T*, at the SNP in exon 7 encoding the amino acid serine and all except two heterozygous (*G/T*) animals, were homozygous for the allele *G* at the SNP in intron 1. Therefore it could well be the case that these less common alleles at these two loci originated from the Meishan breed. It could be speculated that the alleles have been introduced into these European breeds from pigs of Meishan origin. The frequency of these "Meishan" alleles was lowest in the sows of Duroc origin, a breed that was introduced to the UK from Canada and the USA in the early 1980s (http://www.britishpigs.org.uk). However there is no evidence to link the loci tested within *SPP1* to the higher litter sizes seen in the Meishan, controlled through increased prenatal survival levels.

The alleles at the studied SNPs in exon 7 and intron 1 were found to be in strong linkage disequilibrium. Future work could involve the investigation of association of other sites within the *SPP1* gene and litter size in these commercial breeds, after first testing that the loci are not in linkage disequilibrium with these two loci. The control of gene expression operates at many levels and it would be worthwhile searching for recognition sites for gene specific regulatory proteins several kb upstream or

downstream of the promoter. Indeed in higher eucaryotes it is not unusual to find regulatory sequences as far away as 50 kb from the gene they control (Alberts et al., 2002). It would also be worthwhile to utilise the comparative mapping information to search for other physiological candidate genes within the detected QTL region and to investigate these, one specific example being the gene *SPARCL1*.

It was interesting to note that the SINE element in intron 6 appeared to be completely absent within the purebred Meishan animals. Therefore the commercial population of sows could be typed for the presence or absence of the SINE, using the same size separation method as described in section 6.3.1.1, to investigate whether this variant is associated with litter size or more likely whether the presence of this SINE is just an evolutionary artefact. In addition, the putative variants identified for this project were determined from copies of the gene from alternative chromosomes of a single F1 Meishan x Large White boar. Therefore there may well be other variants present in different individuals of Meishan or Large White origin, that were not found within this particular boar.

This project has focussed solely on porcine chromosome 8. There will of course be other loci throughout the porcine genome, which will explain differing amounts of the variation seen in prenatal survival levels between pig breeds. Both embryonic and uterine gene products participate in the control of conceptus growth and development (Simmen et al., 1990). However I will only outline a few possible maternal candidate genes, as it is the regulation by the sow during the peri-implantation period, which is the most critical control of embryo survival in pigs (Roberts et al., 1993). These gene products include uteroferrin, which transports iron to the fetus (Bazer et al., 1991), retinol binding protein (RBP), which transports vital vitamin A to the fetus (Adams et al., 1981), IGF-I and IGF-II, which regulate endometrial growth and differentiation at the time of implantation (Simmen et al., 1990), EGF receptors stimulated by EGF to secrete prostaglandin at the time of implantation (Zhang et al., 1992b) and antileukoproteinase, which maintains the placental cell membrane integrity (Farmer et al., 1990).

In addition, trypsin and chymotrypsin are believed to limit and prevent damage from proteolytic enzymes from the conceptus and may control trophoblast invasiveness (Fazleabas *et al.*, 1982; Fazleabas *et al.*, 1985 and (Mullins et al., 1980). Finally, kallikrein, a serine protease, may be involved with the timing of porcine conceptus expansion and placental attachment to the uterine surface (Vonnahme et al., 1999). All of these factors are secreted in the histotroph (uterine milk) from the endometrium during the peri-implantation period and help control embryo survival, development and successful implantation during this critical period of loss. Using a genomics approach to understand the genetic control of complex traits would require that such physiological candidate genes would only merit further investigation on the basis of positional information i.e. whether they map to a QTL for the trait of interest.

This project has resulted in the production of a gene rich genetic map of porcine chromosome 8 and a detailed comparative map between HSA4 and SSC8. In addition the entire sequence of *SPP1* from Meishan and Large White origin is now known. This gene is a strong physiological as well as positional candidate gene involved in the control in the variation in prenatal survival during the peri-implantation period, seen between these two breeds. However, there was no evidence to suggest that the two SNPs identified at either end of this gene are associated with variation in the resulting trait of litter size within commercial breeding populations.

The markers that define the QTL region around this gene can be used for marker-assisted selection in the meantime, however the identification of the causal locus or loci underlying the genetic variation would provide a more powerful tool for improving the reproductive performance of commercial sows. It is therefore hoped that future investigations will be performed, that will narrow down the candidate region even further than has been possible within this project.

BIBLIOGRAPHY

Adams, K. L., F. W. Bazer, and R. M. Roberts. 1981. Progesterone-Induced Secretion of A Retinol-Binding Protein in the Pig Uterus. J Reprod Fertil 62:39-47.

Adams, M. D., S. E. Celniker, R. A. *et al.* 2000. The genome sequence of Drosophila melanogaster. Science 287:2185-2195.

Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. Mccombie, and J. C. Venter. 1991. Complementary-DNA Sequencing - Expressed Sequence Tags and Human Genome Project. Science 252:1651-1656.

Aggrey, S. E., J. Yao, M. P. Sabour, C. Y. Lin, D. Zadworny, J. F. Hayes, and U. Kuhnlein. 1999. Markers within the regulatory region of the growth hormone receptor gene and their association with milk-related traits in Holsteins. J Hered 90:148-151.

Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2002. Molecular Biology of the Cell. Fourth Edition. Garland Science, Taylor and Francis Group.

Alexander, L. J., G. A. Rohrer, and C. W. Beattie. 1996a. Cloning and characterization of 414 polymorphic porcine microsatellites. Anim Genet 27:137-148.

Alexander, L. J., D. L. Troyer, G. A. Rohrer, T. P. Smith, L. B. Schook, and C. W. Beattie. 1996b. Physical assignments of 68 porcine cosmid and lambda clones containing polymorphic microsatellites. Mamm Genome 7:368-372.

Alfonso, L., J. L. Noguera, D. Babot, and J. Estany. 1997. Estimates of genetic parameters for litter size at different parities in pigs. Livest Prod Sci 47:149-156.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Anderson, L. L. 1974. Pigs. In: E. S. E. Hafez (Ed.) Reproduction in farm animals. Lea and Febiger.

Anderson, S. I., N. L. Lopez-Corrales, B. Gorick, and A. L. Archibald. 2000. A large-fragment porcine genomic library resource in a BAC vector. Mamm Genome 11:811-814.

Andersson, L., A. Archibald, M. Ashburner, S. Audun, W. Barendse, J. Bitgood, C. Bottema, T. Broad, S. Brown, D. Burt, C. Charlier, N. Copeland, S. Davis, M.

Davisson, J. Edwards, A. Eggen, G. Elgar, J. T. Eppig, I. Franklin, P. Grewe, T. Gill, J. A. M. Graves, R. Hawken, J. Hetzel, A. Hilyard, H. Jacob, L. Jaswinska, N. Jenkins, H. Kunz, G. Levan, O. Lie, L. Lyons, P. Maccarone, C. Mellersh, G. Montgomery, S. Moore, C. Moran, D. Morizot, M. Neff, F. Nicholas, S. OBrien, Y. Parsons, J. Peters, J. Postlethwait, M. Raymond, M. Rothschild, L. Schook, Y. Sugimoto, C. Szpirer, M. Tate, J. Taylor, J. VandeBerg, M. Wakefield, J. Wienberg, and J. Womack. 1996. Comparative genome organization of vertebrates. Mamm Genome 7:717-734.

Andersson-Eklund, L., L. Marklund, K. Lundstrom, C. S. Haley, K. Andersson, I. Hansson, M. Moller, and L. Andersson. 1998. Mapping quantitative trait loci for carcass and meat quality traits in a wild boar x large white intercross. J Anim Sci 76:694-700.

Aplin, J. D. 1997. Adhesion molecules in implantation. Rev Reprod 2:84-93.

Archibald, A. L. 1994. Mapping of the pig genome. Curr Opin Genet Dev 4:395-400.

Archibald, A. L. 1998. Comparative genome mapping - the livestock perspective. In: A. J. Clark (Ed.) Animal Breeding - Technology for the 21st Century. pp. 137-164. Harwood academic publishers.

Archibald, A. L., J. F. Brown, C. S. Haley, M. Fredholm, A. K. Wintero, W. Coppieters, A. Van de Weghe, E. Signer, N. J. Larsen, V. H. Nielson, M. Johansson, and L. L. Anderson. 1992. Linkage mapping in the domestic pig (Sus scrofa). Anim Genet Suppl 23:88-89.

Archibald, A. L. and C. S. Haley. 1990. A PiGMaP: An European initiative to map the porcine genome. Anim Genet 22:82-83.

Archibald, A. L. and C. S. Haley. 1998. Genetic linkage maps. In: M. F. Rothschild and A. Ruvinsky (Eds.) The Genetics of the Pig. pp. 265-294. CAB International.

Archibald, A. L., C. S. Haley, J. F. Brown, S. Couperwhite, H. A. McQueen, D. Nicholson, W. Coppieters, A. Van de Weghe, A. Stratil, and A. K. Wintero. 1995. The PiGMaP consortium linkage map of the pig (Sus scrofa). Mamm Genome 6:157-175.

Ashkar, S., L. C. Gerstenfeld, and M. J. Glimcher. 1995. Role of Phosphorylation of Osteopontin in the Formation of Focal Adhesion Between Osteoblasts and the Substratum. J Bone Miner Res 10:S320.

Ashkar, S., D. B. Teplow, M. J. Glimcher, and R. A. Saavedra. 1993. In vitro Phosphorylation of Mouse Osteopontin Expressed in Escherichia-Coli. Biochem Bioph Res Co 191:126-133.

Ashworth, C. J., A. R. Pickard, S. J. Miller, A. P. F. Flint, and J. R. Diehl. 1997. Comparative studies of conceptus-endometrial interactions in Large White x

Landrace and Meishan gilts. Reproduction Fertility and Development 9:217-225.

Avalos, E. and C. Smith. 1987. Genetic-improvement of litter size in pigs. Anim Prod 44:153-164.

Avner, P., T. Bruls, I. Poras, L. Eley, S. Gas, P. Ruiz, M. V. Wiles, R. Sousa-Nunes, R. Kettleborough, A. Rana, J. Morissette, L. Bentley, M. Goldsworthy, A. Haynes, E. Herbert, L. Southam, H. Lehrach, J. Weissenbach, G. Manenti, P. Rodriguez-Tome, R. Beddington, S. Dunwoodie, and R. D. Cox. 2001. A radiation hybrid transcript map of the mouse genome. Nat Genet 29:194-200.

Bayless, K. J., G. A. Meininger, J. M. Scholtz, and G. E. Davis. 1998. Osteopontin is a ligand for the alpha(4)beta(1) integrin. J Cell Sci 111:1165-1174.

Bazer, F. W., D. Worthingtonwhite, M. F. V. Fliss, and S. Gross. 1991. Uteroferrin - A Progesterone-Induced Hematopoietic Growth- Factor of Uterine Origin. Experimental Hematology 19:910-915.

Ben-Dor, A., B. Chor, and D. Pelleg. 2000. RHO - Radiation hybrid ordering. Genome Res 10:365-378.

Benham, F., K. Hart, J. Crolla, M. Bobrow, M. Francavilla, and P. N. Goodfellow. 1989. A Method for Generating Hybrids Containing Nonselected Fragments of Human-Chromosomes. Genomics 4:509-517.

Bennett, G. L. and K. A. Leymaster. 1989. Integration of ovulation rate, potential embryonic viability and uterine capacity into a model of litter size in swine. J Anim Sci 67:1230-1241.

Bernier, F., G. Leblanc, F. Labrie, and V. Luu-The. 1994. Structure of human estrogen and aryl sulfotransferase gene. Two mRNA species issued from a single gene. J Biol Chem 269:28200-28205.

Bertani, G. R., C. D. Gladney, R. K. Johnson, and D. Pomp. Gene expression analysis in anterior pituitary and EST mapping to investigate genetics of swine fertility. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August 19-23. 2002.

Bidanel, J. P., J. C. Caritez, and C. Legault. 1989. Estimation of crossbreeding parameters between Large White and Meishan porcine breeds .1. Reproductive-performance. Genet Sel Evol 21:507-526.

Bidanel, J. P., D. Milan, N. Iannuccelli, Y. Amigues, M. Y. Boscher, F. Bourgeois, J. C. Caritez, J. Gruand, P. Le Roy, H. Lagant, R. Quintanilla, C. Renard, J. Gellin, L. Ollivier, and C. Chevalet. 2001. Detection of quantitative trait loci for growth and fatness in pigs. Genet Sel Evol 33:289-309.

Bidanel, J. P., D. Milan, C. Renard, J. Gruand, and J. Mourot. Detection of quantitative trait loci for intramuscular fat content and lipogenic enzyme

activities in Meishan x Large White F2 pigs. Proceedings of the 7th world congress applied to livestock production. 2002.

Bidanel, J. P. and M. Rothschild. 2002. Current status of quantitative trait locus mapping in pigs. Pig News and Information 23:39-53.

Biensen, N. J., M. E. Wilson, and S. P. Ford. 1998. The impact of either a Meishan or Yorkshire uterus on Meishan or Yorkshire fetal and placental development to days 70, 90, and 110 of gestation. J Anim Sci 76:2169-2176.

Biensen, N. J., M. E. Wilson, and S. P. Ford. 1999. The impacts of uterine environment and fetal genotype on conceptus size and placental vascularity during late gestation in pigs. J Anim Sci 77:954-959.

Bihoreau, M. T., L. Sebag-Montefiore, R. F. Godfrey, R. H. Wallis, J. H. Brown, P. A. Danoy, S. C. Collins, M. Rouard, P. J. Kaisaki, M. Lathrop, and D. Gauguier. 2001. A high-resolution consensus linkage map of the rat, integrating radiation hybrid and genetic maps. Genomics 75:57-69.

Bishop, M. Comparative mapping in man and vertebrates. http://www.hgmp.mrc.ac.uk (UK Human Genome Mapping Project website). 2000.

Blattman, A. N., B. W. Kirkpatrick, and K. E. Gregory. 1996. A search for quantitative trait loci for ovulation rate in cattle. Anim Genet 27:157-162.

Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. ColladoVides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of Escherichia coli K-12. Science 277:1453-1462.

Boehnke, M., K. Lange, and D. R. Cox. 1991. Statistical methods for multipoint radiation hybrid mapping. Am J Hum Genet 49:1174-1188.

Bonfield, J. K., K. F. Beal, M. J. Betts, and R. Staden. 2002. Trev: a DNA trace editor and viewer. Bioinformatics 18:194-195.

Bowen, J. A., F. W. Bazer, and R. C. Burghardt. 1997. Spatial and temporal analyses of integrin and Muc-1 expression in porcine uterine epithelium and trophectoderm in vitro. Biol Reprod 56:409-415.

Braunschweig, M. H., A. A. Paszek, J. I. Weller, Y. Da, R. J. Hawken, M. B. Wheeler, L. B. Schook, and L. J. Alexander. 2001. Generation and exploration of a dense genetic map in a region of a QTL affecting corpora lutea in a Meishan x Yorkshire cross. Mamm Genome 12:719-723.

Burt, D. W. 2002. Origin and evolution of avian microchromosomes. Cytogenetic and Genome Research 96:97-112.

Burt, D. W., C. Bruley, I. C. Dunn, C. T. Jones, A. Ramage, A. S. Law, D. R. Morrice, I. R. Paton, J. Smith, D. Windsor, A. Sazanov, R. Fries, and D. Waddington. 1999. The dynamics of chromosome evolution in birds and mammals. Nature 402:411-413.

Cameron, N. D., M. K. Curran, and R. Thompson. 1988. Estimation of sire with feeding regime interaction in pigs. Anim Prod 46:87-95.

Campbell, E. M. and G. A. Rohrer. A polymorphism in the putative start codon of alpha mannosidase 2B2 does not appear to affect ovulation rate in swine. Proceedings of the XXVIIth International Conference on Animal Genetics, Minneapolis, USA, July 22-26. 2000.

Cardon, L. R. and J. I. Bell. 2001. Association study designs for complex diseases. Nat Rev Genet 2:91-99.

Cargill, E. J., L. C. Baskin, and D. Pomp. 1998. Localization of the porcine Carboxypeptidase-E gene by linkage analysis further extends the region of synteny between human chromosome 4 and porcine chromosome 8. J Anim Sci 76:2211-2212.

Casas, E., J. W. Keele, S. D. Shackelford, M. Koohmaraie, T. S. Sonstegard, T. P. L. Smith, S. M. Kappes, and R. T. Stone. 1998. Association of the muscle hypertrophy locus with carcass traits in beef cattle. J Anim Sci 76:468-473.

Cassady, J. P., R. K. Johnson, D. Pomp, G. A. Rohrer, L. D. Van Vleck, E. K. Spiegel, and K. M. Gilson. 2001. Identification of quantitative trait loci affecting reproduction in pigs. J Anim Sci 79:623-633.

Chakravarti, A. 2001. Single nucleotide polymorphisms ... to a future of genetic medicine. Nature 409:822-823.

Charlier, C., W. Coppieters, F. Farnir, L. Grobet, P. L. Leroy, C. Michaux, M. Mni, A. Schwers, P. Vanmanshoven, R. Hanset, and M. Georges. 1995. The Mh Gene Causing Double-Muscling in Cattle Maps to Bovine Chromosome-2. Mamm Genome 6:788-792.

Cheng, P. L. 1983. A highly prolific pig breed in China - the Taihu pig. Pig News and Information 4:407-425.

Cheung, V. G. and R. S. Spielman. 2002. The genetics of variation in gene expression. Nat Genet Suppl 32:522-525.

Chowdhary, B. P. 1998. Cytogenetics and physical chromsome maps. In: M. F. Rothschild and A. Ruvinsky (Eds.) The Genetics of the Pig. pp. 199-264. CAB International.

Churchill, G. A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Cockett, N. E., S. P. Jackson, T. L. Shay, D. Nielsen, S. S. Moore, M. R. Steele, W. Barendse, R. D. Green, and M. Georges. 1994. Chromosomal Localization of the Callipyge Gene in Sheep (Ovis- Aries) Using Bovine Dna Markers. Proc Natl Acad Sci U S A 91:3019-3023.

Coghlan, A. and P. Cohen. 2002. Of mice and men. New Scientist 7 December 2002:12-13.

Collins, F. S., M. S. Guyer, and A. Chakravarti. 1997. Variations on a theme: Cataloging human DNA sequence variation. Science 278:1580-1581.

Collins, F. S., A. Patrinos, E. Jordan, A. Chakravarti, R. Gesteland, and L. Walters. 1998. New Goals for the U.S. Human Genome Project: 1998-2003. Science 282:682-689.

Copeland, N. G., N. A. Jenkins, and D. Court. 2001. Recombineering: A powerful new tool for mouse functional genomics. Nat Rev Genet 2:769-779.

Coppieters, W., S. Blott, F. Farnir, B. Grisart, J. Riquet, and M. Georges. 1999. From phenotype to genotype: towards positional cloning of QTL in livestock? Archiv fur Tierzucht-Archives of Animal Breeding 42:86-92.

Coppieters, W., J. Riquet, J. J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 1998. A QTL with major effect on milk yield and composition maps to bovine Chromosome 14. Mamm Genome 9:540-544.

Coppieters, W., A. Vandeweghe, L. Peelman, A. Depicker, A. Vanzeveren, and Y. Bouquet. 1993. Characterization of porcine polymorphic microsatellite loci. Anim Genet 24:163-170.

Corpet, F. 1988. Multiple Sequence Alignment with Hierarchical-Clustering. Nucleic Acids Res 16:10881-10890.

Couzin, J. 2002. Genomics - New mapping project splits the community. Science 296:1391-1393.

Cox, D. R., M. Burmeister, E. R. Price, S. Kim, and R. M. Myers. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250:245-250.

Craig, A. M. and D. T. Denhardt. 1991. The Murine Gene Encoding Secreted Phosphoprotein-1 (Osteopontin) - Promoter Structure, Activity, and Induction Invivo by Estrogen and Progesterone. Gene 100:163-171.

Davies, W., I. Harbitz, R. Fries, G. Stranzinger, and J. G. Hauge. 1988. Porcine Malignant Hyperthermia Carrier Detection and Chromosomal Assignment Using A Linked Probe. Anim Genet 19:203-212.

Dawson, E., G. R. Abecasis, S. Bumpstead, Y. Chen, S. Hunt, D. M. Beare, J. Pabial, T. Dibling, E. Tinsley, S. Kirby, D. Carter, M. Papaspyridonos, S. Livingstone, R. Ganske, E. Lohmmussaar, J. Zernant, N. Tonisson, M. Remm, R. Magi, T. Puurand, J. Vilo, A. Kurg, K. Rice, P. Deloukas, R. Mott, A. Metspalu, D. R. Bentley, L. R. Cardon, and I. Dunham. 2002. A first-generation linkage disequilibrium map of human chromosome 22. Nature 418:544-548.

De Koning, D. J., A. P. Rattink, B. Harlizius, M. A. M. Groenen, E. W. Brascamp, and J. A. M. van Arendonk. 2001. Detection and characterization of quantitative trait loci for growth and reproduction traits in pigs. Livest Prod Sci 72:185-198.

Denhardt, D. T. 1966. A membrane-filter technique for the detection of complementary DNA. Biochem Bioph Res Co 23:641-646.

Denhardt, D. T. and M. Noda. 1998. Osteopontin expression and function: Role in bone remodeling. J Cell Biochem Suppl 30/31:92-102.

Desautes, C., J. P. Bidanel, D. Milan, N. Iannuccelli, Y. Amigues, F. Bourgeois, J. C. Caritez, C. Renard, C. Chevalet, and P. Mormede. 2002. Genetic linkage mapping of quantitative trait loci for behavioral and neuroendocrine stress response traits in pigs. J Anim Sci 80:2276-2285.

Drogemuller, C., H. Hamann, and O. Distl. 2001. Candidate gene markers for litter size in different German pig lines. J Anim Sci 79:2565-2570.

Dube, J. L., P. Wang, J. Elvin, K. M. Lyons, A. J. Celeste, and M. M. Matzuk. 1998. The bone morphogenetic protein 15 gene is X-linked and expressed in oocytes. Mol Endocrinol 12:1809-1817.

Echard, G., T. E. Broad, D. Hill, and P. Pearce. 1994. Present status of the ovine gene map (Ovis aries); comparison with the bovine map (Bos taurus). Mamm Genome 5:324-332.

Edfors-Lilja, I., E. Wattrang, L. Marklund, M. Moller, L. Andersson-Eklund, L. Andersson, and C. Fossum. 1998. Mapping quantitative trait loci for immune capacity in the pig. Journal of Immunology 161:829-835.

Ellegren, H., B. P. Chowdhary, M. Johansson, L. Marklund, M. Fredholm, I. Gustavsson, and L. Andersson. 1994. A primary linkage map of the porcine genome reveals a low-rate of genetic recombination. Genetics 137:1089-1100.

Ellegren, H., M. Fredholm, I. Edfors-Lilja, A. K. Wintero, and L. Andersson. 1993a. Conserved synteny between pig chromosome 8 and human chromosome 4 but rearranged and distorted linkage maps. Genomics 17:599-603.

Ellegren, H., M. Johansson, B. P. Chowdhary, S. Marklund, D. Ruyter, L. Marklund, P. Braunernielsen, I. Edfors-Lilja, I. Gustavsson, R. K. Juneja, and L.

Andersson. 1993b. Assignment of 20 microsatellite markers to the porcine linkage map. Genomics 16:431-439.

Ewen, K. R., G. S. A. Myers, L. M. Dube, J. W. Barlow, S. J. Foote, and K. Poetter. A SNP genotyping method using microarray technology. Proceedings of the Human Genome meeting, Edinburgh, UK, April 19-22, 2001.

Excoffier, L. and M. Slatkin. 1995. Maximum-Likelihood-Estimation of Molecular Haplotype Frequencies in A Diploid Population. Mol Biol Evol 12:921-927.

Fahrenkrug, S. C., B. A. Freking, T. P. L. Smith, G. A. Rohrer, and J. W. Keele. 2002. Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. Anim Genet 33:186-195.

Fan, J. B., X. Q. Chen, M. K. Halushka, A. Berno, X. H. Huang, T. Ryder, R. J. Lipshutz, D. J. Lockhart, and A. Chakravarti. 2000. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. Genome Res 10:853-860.

Farmer, S. J., A. E. Fliss, and R. C. M. Simmen. 1990. Complementary-Dna Cloning and Regulation of Expression of the Messenger-Rna Encoding A Pregnancy-Associated Porcine Uterine Protein Related to Human Antileukoproteinase. Mol Endocrinol 4:1095-1104.

Farnir, F., B. Grisart, W. Coppieters, J. Riquet, P. Berzi, N. Cambisano, L. Karim, M. Mni, S. Moisio, P. Simon, D. Wagenaar, J. Vilkki, and M. Georges. 2002. Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: Revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. Genetics 161:275-287.

Fazleabas, A. T., F. W. Bazer, P. J. Hansen, R. D. Geisert, and R. M. Roberts. 1985. Differential Patterns of Secretory Protein Localization Within the Pig Uterine Endometrium. Endocrinology 116:240-245.

Fazleabas, A. T., F. W. Bazer, and R. M. Roberts. 1982. Purification and Properties of A Progesterone-Induced Plasmin Trypsin-Inhibitor from Uterine Secretions of Pigs and Its Immuno-Cytochemical Localization in the Pregnant Uterus. J Biol Chem 257:6886-6897.

Fazleabas, A. T., S. C. Bell, S. Fleming, J. H. Sun, and B. A. Lessey. 1997. Distribution of integrins and the extracellular matrix proteins in the baboon endometrium during the menstrual cycle and early pregnancy. Biol Reprod 56:348-356.

Feinberg, A. P. and B. Vogelstein. 1983. A Technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity. Anal Biochem 132:6-13.

Feinberg, A. P. and B. Vogelstein. 1984. A Technique for Radiolabeling DNA Restriction Endonuclease Fragments to High Specific Activity. Addendum. Anal Biochem 137:266-267.

Fisher, R. A. 1935. The Design of Experiments. Oliver and Boyd Ltd., London.

Fisher, R. A. 1958. The Genetical Theory of Natural Selection. Dover, New York.

Fong, A. M., L. A. Robinson, D. A. Steeber, T. F. Tedder, O. Yoshie, T. Imai, and D. D. Patel. 1998. Fractalkine and CX3CR1 mediate a novel mechanism of leukocyte capture, firm adhesion, and activation under physiologic flow. J Exp Med 188:1413-1419.

Ford, S. P. 1997. Embryonic and fetal development in different genotypes in pigs. J Reprod Fertil Suppl 165-176.

Fredholm, M., A. K. Wintero, K. Christensen, B. Kristensen, P. B. Nielsen, W. Davies, and A. L. Archibald. 1993. Characterization of 24 porcine (dA-dC)n-(dT-dG)n microsatellites - genotyping of unrelated animals from 4 breeds and linkage studies. Mamm Genome 4:187-192.

Freking, B. A., S. K. Murphy, A. A. Wylie, S. J. Rhodes, J. W. Keele, K. A. Leymaster, R. L. Jirtle, and T. P. L. Smith. 2002. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome Res 12:1496-1506.

Fujii, J., K. Otsu, F. Zorzato, S. Deleon, V. K. Khanna, J. E. Weiler, P. J. Obrien, and D. H. Maclennan. 1991. Identification of A Mutation in Porcine Ryanodine Receptor Associated with Malignant Hyperthermia. Science 253:448-451.

Galloway, S. M., K. P. McNatty, L. M. Cambridge, M. P. E. Laitinen, J. L. Juengel, T. S. Jokiranta, R. J. McLaren, K. Luiro, K. G. Dodds, G. W. Montgomery, A. E. Beattie, G. H. Davis, and O. Ritvos. 2000. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. Nat Genet 25:279-283.

Galvin, J. M., I. Wilmut, B. N. Day, M. Ritchie, M. Thomson, and C. S. Haley. 1993. Reproductive performance in relation to uterine and embryonic traits during early gestation in Meishan, large white and crossbred sows. J Reprod Fertil 98:377-384.

Garlow, J. E., H. Ka, G. A. Johnson, R. C. Burghardt, L. A. Jaeger, and F. W. Bazer. 2002. Analysis of osteopontin at the maternal-placental interface in pigs. Biol Reprod 66:718-725.

Gavora, J. S., U. Kuhnlein, L. B. Crittenden, J. L. Spencer, and M. P. Sabour. 1991. Endogenous Viral Genes - Association with Reduced Egg- Production Rate and Egg Size in White Leghorns. Poultry Science 70:618-623.

Gavrilets, S. 2001. Population genetics: Multilocus. In: Encyclopedia of life sciences. Macmillian Publishers Ltd.

Geisert, R. D. 1997. Porcine implantation: Regulation of uterine receptivity to trophoblast attachment while preventing endometrial, penetration by the invasive conceptus. Biol Reprod 56:M17.

Geisert, R. D., R. H. Renegar, W. W. Thatcher, R. M. Roberts, and F. W. Bazer. 1982. Establishment of Pregnancy in the Pig .1. Interrelationships Between Pre-Implantation Development of the Pig Blastocyst and Uterine Endometrial Secretions. Biol Reprod 27:925-939.

Geisert, R. D. and J. V. Yelich. 1997. Regulation of conceptus development and attachment in pigs. J Reprod Fertil133-149.

Georges, M. 1999. Towards marker assisted selection in livestock. Reprod Nutr Dev 39:555-561.

Gerdes, D., M. Wehling, B. Leube, and E. Falkenstein. 1998. Cloning and tissue expression of two putative steroid membrane receptors. Biol Chem 379:907-911.

Gibson, J. P., Z. H. Jiang, J. A. B. Robinson, A. L. Archibald, and C. S. Haley. 2002. No detectable association of the ESR PvuII mutation with sow productivity in a Meishan x Large White F-2 population. Anim Genet 33:448-450.

Gillard, E. F., K. Otsu, J. Fujii, V. K. Khanna, S. Deleon, J. Derdemezi, B. A. Britt, C. L. Duff, R. G. Worton, and D. H. Maclennan. 1991. A substitution of cysteine for arginine-614 in the ryanodine receptor is potentially causative of human-malignant hyperthermia. Genomics 11:751-755.

Girard, J. P. and T. A. Springer. 1995. Cloning from Purified High Endothelial Venule Cells of Hevin, A Close Relative of the Antiadhesive Extracellular-Matrix Protein Sparc. Immunity 2:113-123.

Goss, S. J. and H. Harris. 1975. New method for mapping genes in human chromosomes. Nature 255:680-684.

Green, E. D., R. M. Mohr, J. R. Idol, M. Jones, J. M. Buckingham, L. L. Deaven, R. K. Moyzis, and M. V. Olson. 1991. Systematic Generation of Sequence-Tagged Sites for Physical Mapping of Human-Chromosomes - Application to the Mapping of Human Chromosome-7 Using Yeast Artificial Chromosomes. Genomics 11:548-564.

Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res 12:222-231.

Gut, I. G. 2001. Automation in genotyping of single nucleotide polymorphisms. Human Mutation 17:475-492.

Gyapay, G., K. Schmitt, C. Fizames, H. Jones, N. Vega-Czarny, D. Spillett, D. Muselet, J. F. Prud'Homme, C. Dib, C. Auffray, J. Morissette, J. Weissenbach, and P. N. Goodfellow. 1996. A radiation hybrid map of the human genome. Hum Mol Genet 5:339-346.

Haley, C. S. and L. Andersson. 1997. Linkage mapping of quantitative trait loci in plants and animals. In: P. H. Dear (Ed.) Genome mapping - a practical approach. pp. 49-71. Oxford University Press.

Haley, C. S. and A. L. Archibald. 1992. Porcine genome analysis. In: K. E. Davies and S. M. Tilghman (Eds.) Strategies for physical mapping (volume 4). pp. 99-130. Cold spring harbor laboratory press.

Haley, C. S. and A. L. Archibald. The Pig Gene Mapping Project (PiGMaP). (http://www.ri.bbsrc.ac.uk/pigmap/book.html). 1999.

Haley, C. S., E. D'agaro, and M. Ellis. 1992. Genetic components of growth and ultrasonic fat depth traits in Meishan and Large White pigs and their reciprocal crosses. Anim Prod 54:105-115.

Haley, C. S., S. A. Knott, and J. M. Elsen. 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136:1195-1207.

Haley, C. S. and G. J. Lee. 1993. Genetic basis of prolificacy in Meishan pigs. J Reprod Fertil Suppl 48:247-259.

Haley, C. S., G. J. Lee, and M. Ritchie. 1995. Comparative reproductive-performance in Meishan and Large White pigs and their crosses. Anim Sci 60:259-267.

Hanenberg, E. H. A. T., E. F. Knol, and J. W. M. Merks. 2001. Estimates of genetic parameters for reproduction traits at different parities in Dutch Landrace pigs. Livest Prod Sci 69:179-186.

Hanset, R., C. Dasnois, S. Scalais, C. Michaux, and L. Grobet. 1995. Introgression Into the Pietrain Genome of the Normal Allele at the Locus for Halothane Sensitivity. Genet Sel Evol 27:77-88.

Hawken, R. J., J. Murtaugh, G. H. Flickinger, M. Yerle, A. Robic, D. Milan, J. Gellin, C. W. Beattie, L. B. Schook, and L. J. Alexander. 1999. A first-generation porcine whole-genome radiation hybrid map. Mamm Genome 10:824-830.

Hediger, R., H. A. Ansari, and G. F. Stranzinger. 1991. Chromosome-Banding and Gene Localizations Support Extensive Conservation of Chromosome Structure Between Cattle and Sheep. Cytogenet Cell Genet 57:127-134.

Her, C., I. A. Aksoy, S. Kimura, B. F. Brandriff, J. J. Wasmuth, and R. M. Weinshilboum. 1995. Human estrogen sulfotransferase gene (STE): cloning, structure, and chromosomal localization. Genomics 29:16-23.

Heyen, D. W., J. I. Weller, M. Ron, M. Band, J. E. Beever, E. Feldmesser, Y. Da, G. R. Wiggans, P. M. VanRaden, and H. A. Lewin. 1999. A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol Genomics 1:165-175.

Hijiya, N., M. Setoguchi, K. Matsuura, Y. Higuchi, S. Akizuki, and S. Yamamoto. 1994. Cloning and characterisation of the human osteopontin gene and its promoter. Biochem J 303:255-262.

Hirooka, H., D. J. De Koning, B. Harlizius, J. A. M. van Arendonk, A. P. Rattink, M. A. M. Groenen, E. W. Brascamp, and H. Bovenhuis. 2001. A whole-genome scan for quantitative trait loci affecting teat number in pigs. J Anim Sci 79:2320-2326.

Hu, D. D., E. C. K. Lin, N. L. Kovach, J. R. Hoyer, and J. W. Smith. 1995. A biochemical-characterization of the binding of osteopontin to integrins alpha(V)beta(1) and alpha(V)beta(5). J Biol Chem 270:26232-26238.

Hu, J., C. Mungall, A. Law, R. Papworth, J. P. Nelson, A. Brown, I. Simpson, S. Leckie, D. W. Burt, A. L. Hillyard, and A. L. Archibald. 2001. The ARKdb: genome databases for farmed and other animals. Nucleic Acids Res 29:106-110.

Hudson, T. J., D. M. Church, S. Greenaway, H. Nguyen, A. Cook, R. G. Steen, W. J. Van Etten, A. B. Castle, M. A. Strivens, P. Trickett, C. Heuston, C. Davison, A. Southwell, R. Hardisty, A. Varela-Carver, A. R. Haynes, P. Rodriguez-Tome, H. Doi, M. S. H. Ko, J. Pontius, L. Schriml, L. Wagner, D. Maglott, S. D. M. Brown, E. S. Lander, G. Schuler, and P. Denny. 2001. A radiation hybrid map of mouse genes. Nat Genet 29:201-205.

Hynes, R. O. 1996. Targeted mutations in cell adhesion genes: What have we learned from them? Dev Biol 180:402-412.

Innes, M. A. and D. H. Gelfand. 1990. Optimisation of PCRs. In: M. A. Innes, D. H. Gelfand, J. J. Sninsky, and T. White (Eds.) PCR protocols: a guide to methods and applications. pp. 3-12. Academic Press.

Itoh, A., T. Miyabayashi, M. Ohno, and S. Sakano. 1998. Cloning and expressions of three mammalian homologues of Drosophila slit suggest possible roles for Slit in the formation and maintenance of the nervous system. Mol Brain Res 62:175-186.

Iwasaki, H., Y. Ezura, R. Ishida, M. Kajita, M. Kodaira, J. Knight, S. Daniel, M. Shi, and M. Emi. 2002. Accuracy of genotyping for single nucleotide polymorphisms by a microarray-based single nucleotide polymorphism

typing method involving hybridization of short allele-specific oligonucleotides. DNA Res 9:59-62.

Janke, A., G. Feldmaierfuchs, W. K. Thomas, A. Vonhaeseler, and S. Paabo. 1994. The Marsupial Mitochondrial Genome and the Evolution of Placental Mammals. Genetics 137:243-256.

Janss, L. L. G., J. A. M. VanArendonk, and E. W. Brascamp. 1997. Segregation analyses for presence of major genes affecting growth, backfat, and litter size in Dutch Meishan crossbreds. J Anim Sci 75:2864-2876.

Jiang, Z., C. Priat, and F. Galibert. 1998. Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. Mamm Genome 9:577-587.

Jiang, Z. H., J. P. Gibson, A. L. Archibald, and C. S. Haley. 2001. The porcine gonadotropin-releasing hormone receptor gene (GNRHR): Genomic organization, polymorphisms, and association with the number of corpora lutea. Genome 44:7-12.

Jiang, Z. H., H. He, N. Hamasima, H. Suzuki, and A. M. V. Gibbins. 2002a. Comparative mapping of Homo sapiens chromosome 4 (HSA4) and Sus scrofa chromosome 8 (SSC8) using orthologous genes representing different cytogenetic bands as landmarks. Genome 45:147-156.

Jiang, Z. H., J. A. B. Robinson, A. M. Verrinder Gibbins, J. P. Gibson, A. L. Archibald, and C. S. Haley. Mapping of QTL for prolificacy traits on SSC8 using a candidate gene approach. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France, August 19-23, 2002b.

Johansson, M., H. Ellegren, and L. Andersson. 1995. Comparative mapping reveals extensive linkage conservation - but with gene order rearrangements - between the pig and the human genomes. Genomics 25:682-690.

Johnson, G. A., R. C. Burghardt, T. E. Spencer, G. C. R. Newton, T. L. Ott, and F. W. Bazer. 1999a. Ovine osteopontin: II. Osteopontin and alpha(v)beta(3) integrin expression in the uterus and conceptus during the periimplantation period. Biol Reprod 61:892-899.

Johnson, G. A., T. E. Spencer, R. C. Burghardt, and F. W. Bazer. 1999b. Ovine osteopontin: I. Cloning and expression of messenger ribonucleic acid in the uterus during the periimplantation period. Biol Reprod 61:884-891.

Johnson, R. K., M. K. Nielsen, and D. S. Casey. 1999c. Responses in ovulation rate, embryonal survival, and litter traits in swine to 14 generations of selection to increase litter size. J Anim Sci 77:541-557.

Johnson, G. A., T. E. Spencer, R. C. Burghardt, K. M. Taylor, C. A. Gray, and F. W. Bazer. 2000. Progesterone modulation of osteopontin gene expression in the ovine uterus. Biol Reprod 62:1315-1321.

Johnson, G. A., F. W. Bazer, L. A. Jaeger, H. Ka, J. E. Garlow, C. Pfarrer, T. E. Spencer, and R. C. Burghardt. 2001. Muc-1, integrin, and osteopontin expression during the implantation cascade in sheep. Biol Reprod 65:820-828.

Johnson, M. H. and B. J. Everitt. 1995. Essential Reproduction. Blackwell Science.

Jonmundsson, J. V. and S. Adalsteinsson. 1985. Single genes for fecundity in Icelandic sheep. In: R. B. Land and D. W. Robinson (Eds.) Genetics of reproduction in sheep. pp. 159-168. Butterworths, London.

Jung, M., Y. Chen, and H. Geldermann. 1994. Nine porcine polymorphic microsatellites (S0141-S0149). Anim Genet 25:378.

Kambadur, R., M. Sharma, T. P. L. Smith, and J. J. Bass. 1997. Mutations in myostatin (GDF8) in double-muscled Belgian blue and Piedmontese cattle. Genome Res 7:910-916.

Kappes, S. M., G. L. Bennett, J. W. Keele, S. E. Echternkamp, K. E. Gregory, and R. M. Thallman. 2000. Initial results of genomic scans for ovulation rate in a cattle population selected for increased twinning rate. J Anim Sci 78:3053-3059.

Kappes, S. M., J. W. Keele, R. T. Stone, T. S. Sonstegard, T. P. L. Smith, R. A. Mcgraw, N. L. LopezCorrales, and C. W. Beattie. 1997. A second-generation linkage map of the bovine genome. Genome Res 7:235-249.

Kartsogiannis, V., H. Zhou, N. J. Horwood, R. J. Thomas, D. K. Hards, J. M. Quinn, P. Niforas, K. W. Ng, T. J. Martin, and M. T. Gillespie. 1999. Localization of RANKL (receptor activator of NF kappa B ligand) mRNA and protein in skeletal and extraskeletal tissues. Bone 25:525-534.

Keightley, P. D. and S. A. Knott. 1999. Testing the correspondence between map positions of quantitative trait loci. Genet Res 74:323-328.

Kennes, Y. M., B. D. Murphy, F. Pothier, and M. F. Palin. 2001. Characterization of swine leptin (LEP) polymorphisms and their association with production traits. Anim Genet 32:215-218.

Kerr, J. C. and N. D. Cameron. 1996. Genetic and phenotypic relationships between performance test and reproduction traits in Large White pigs. Anim Sci 62:531-540.

Kerr, J. M., L. W. Fisher, J. D. Termine, and M. F. Young. 1991. The cDNA cloning and RNA distribution of bovine osteopontin. Gene 108:237-243.

Kim, J. G., J. H. Song, J. L. Vallet, G. A. Rohrer, G. A. Johnson, M. M. Joyce, and R. K. Christenson. Molecular characterisation and expression of porcine bone morphogenetic protein receptor-1B in the uterus of cylic and pregnant gilts. Biology of Reproduction . 2003.

King, R. D. and M. J. E. Sternberg. 1996. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. Protein Sci 5:2298-2310.

Kirkpatrick, B. W., B. M. Byla, and K. E. Gregory. 2000. Mapping quantitative trait loci for bovine ovulation rate. Mamm Genome 11:136-139.

Kirkpatrick, B. W., A. Mengelt, N. Schulman, and I. C. A. Martin. 1998. Identification of quantitative trait loci for prolificacy and growth in mice. Mamm Genome 9:97-102.

Kmiec, M., A. Dybus, and A. Terman. 2001. Prolactin receptor gene polymorphism and its association with litter size in Polish Landrace. Archiv fur Tierzucht-Archives of Animal Breeding 44:547-551.

Knoll, A., A. Stratil, S. Cepica, and J. Dvorak. 1999. Length polymorphism in an intron of the porcine osteopontin (SPP1) gene is caused by the presence or absence of a SINE (PRE-1) element. Anim Genet 30:466.

Knott, S. A., L. Marklund, C. S. Haley, K. Andersson, W. Davies, H. Ellegren, M. Fredholm, I. Hansson, B. Hoyheim, K. Lundstrom, M. Moller, and L. Andersson. 1998. Multiple marker mapping of quantitative trait loci in a cross between outbred wild boar and large white pigs. Genetics 149:1069-1080.

Korwin-Kossakowska, A., M. Kamyczek, D. Cieslak, M. Pierzchala, and J. Kuryl. The estimation of the candidate genes polymorphism effect on the reproductive traits in line 990 sows. Proceedings of the XXVIIIth International Conference on Animal Genetics, Göttingen, Germany, August 11-15, 2002.

Krebsbach, P. H., S. K. Lee, Y. Matsuki, C. A. Kozak, K. M. Yamada, and Y. Yamada. 1996. Full-length sequence, localization, and chromosomal mapping of ameloblastin. A novel tooth-specific gene. J Biol Chem 271:4431-4435.

Kreiss, T. and R. Vale. 1993. Guidebook to the extracellular matrix and adhesion proteins. Oxford University Press.

Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139-144.

Kumlien, J., A. Grigoriev, C. H. Roest, M. Ross, P. N. Goodfellow, and H. Lehrach. 1996. A radiation hybrid map spanning the entire human X chromosome integrating YACs, genes, and STS markers. Mamm Genome 7:758-766.

Lande, R. and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743-756.

Lander, E. and L. Kruglyak. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241-247.

Lander, E. S., L. M. Linton, B. *et al.* 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Larsen, N. J., S. Marklund, K. A. Kelly, M. Malek, C. K. Tuggle, M. Yerle, and M. F. Rothschild. 1999. New insights into porcine-human synteny conservation. Mamm Genome 10:488-491.

Lee, G. J. and C. S. Haley. 1995. Comparative farrowing to weaning performance in Meishan and Large White pigs and their crosses. Anim Sci 60:269-280.

Lee, G. J., M. Ritchie, M. Thomson, A. A. Macdonald, A. Blasco, M. A. Santacreu, M. J. Argente, and C. S. Haley. 1995. Uterine capacity and prenatal survival in Meishan and Large White pigs. Anim Sci 60:471-479.

Lessey, B. A., I. T. Yeh, A. J. Castelbaum, M. A. Fritz, A. O. Ilesanmi, P. Korzeniowski, J. H. Sun, and K. Chwalisz. 1996. Endometrial progesterone receptors and markers of uterine receptivity in the window of implantation. Fertil Steril 65:477-483.

Li, W. H., M. Gouy, P. M. Sharp, C. O. Uigin, and Y. W. Yang. 1990. Molecular Phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and Molecular Clocks. Proc Natl Acad Sci U S A 87:6703-6707.

Liaw, L., D. E. Birk, G. B. Ballas, J. S. Whitsitt, J. M. Davidson, and B. L. M. Hogan. 1998. Altered wound healing in mice lacking a functional osteopontin gene (spp1). J Clin Invest 101:1468-1478.

Lien, S., A. Karlsen, G. Klemetsdal, D. I. Vage, I. Olsaker, H. Klungland, M. Aasland, B. Heringstad, J. Ruane, and L. Gomez-Raya. 2000. A primary screen of the bovine genome for quantitative trait loci affecting twinning rate. Mamm Genome 11:877-882.

Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Laviolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J. B. Fan, T. Gingeras, J. Warrington, N. Patil, T. J. Hudson, and E. S. Lander. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nat Genet 24:381-386.

Liu, Q., E. C. Thorland, J. A. Heit, and S. S. Sommer. 1997. Overlapping PCR for bidirectional PCR amplification of specific alleles: A rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. Genome Res 7:389-398.

Looft, C., N. Reinsch, C. Karall-Albrecht, S. Paul, M. Brink, H. Thomsen, G. Brockmann, C. Kuhn, M. Schwerin, and E. Kalm. 2001. A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine Chromosome 14. Mamm Genome 12:646-650.

Lunetta, K. L. and M. Boehnke. 1994. Multipoint radiation hybrid mapping: comparison of methods, sample size requirements, and optimal study characteristics. Genomics 21:92-103.

Lunetta, K. L., M. Boehnke, K. Lange, and D. R. Cox. 1995. Experimental-design and error-detection for polyploid radiation hybrid mapping. Genome Res 5:151-163.

Lunetta, K. L., M. Boehnke, K. Lange, and D. R. Cox. 1996. Selected locus and multiple panel models for radiation hybrid mapping. Am J Hum Genet 59:717-725.

Lynch, M. and B. Walsh. 1998. Mapping QTLs: Outbred populations. In: Genetics and Analysis of Quantitative Traits. pp. 491-532. Sinauer Associates Inc.

Lyons, L. A., M. M. Raymond, and S. J. O'Brien. 1994. Comparative genomics: the next generation. Anim Biotechnol 5:103-111.

Maak, S., S. Jaesert, K. Neumann, M. Yerle, and G. von Lengerken. 2001. Isolation of expressed sequence tags of skeletal muscle of neonatal healthy and splay leg piglets and mapping by somatic cell hybrid analysis. Anim Genet 32:303-307.

Maarek, Y. S., M. Jacovi, M. Shtalhaim, S. Ur, D. Zernik, and I. Z. BenShaul. 1997. WebCutter: A system for dynamic and tailorable site mapping. Computer Networks and ISDN Systems 29:1269-1279.

MacDougall, M., B. R. Dupont, D. Simmons, B. Reus, P. Krebsbach, C. Karrman, G. Holmgren, R. J. Leach, and K. Forsman. 1997. Ameloblastin gene (AMBN) maps within the critical region for autosomal dominant amelogenesis imperfecta at chromosome 4q2. Genomics 41:115-118.

Malek, M., J. C. M. Dekkers, H. K. Lee, T. J. Baas, and M. F. Rothschild. 2001. A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. Mamm Genome 12:630-636.

Mandiki, S. N. M., B. Noel, J. L. Bister, R. Peeters, G. Beerlandt, E. Decuypere, A. Visscher, R. Suess, K. H. Kaulfuss, and R. Paquay. 2000. Pre-ovulatory follicular characteristics and ovulation rates in different breed crosses, carriers or non-carriers of the Booroola or Cambridge fecundity gene. Anim Reprod Sci 63:77-88.

Marth, G. T., I. Korf, M. D. Yandell, R. T. Yeh, Z. J. Gu, H. Zakeri, N. O. Stitziel, L. Hillier, P. Y. Kwok, and W. R. Gish. 1999. A general approach to single-nucleotide polymorphism discovery. Nat Genet 23:452-456.

Matise, T. C., M. Perlin, and A. Chakravarti. 1994. Automated construction of genetic-linkage maps using an expert- system (Multimap) - A human genome linkage map. Nat Genet 6:384-390.

McCarthy, L. C. 1996. Whole genome radiation hybrid mapping. Trends Genet 12:491-493.

McEwan, J. C., K. A. Paterson, A. Zadissa, T. van Stijn, C. Diez-Tascon, and A. M. Crawford. TIPS: A process for rapid fine mapping of QTLs using ESTs and comparative mapping. Proc Assoc Advmt Anim Breed Genet 14, 103-106. 2002.

McPherron, A. C., A. M. Lawler, and S. J. Lee. 1997. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. Nature 387:83-90.

McPherson, J. D., M. Marra, L. Hillier, R. H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E. R. Mardis, R. K. Wilson, R. Fulton, T. A. Kucaba, C. Wagner-McPherson, W. B. Barbazuk, S. G. Gregory, S. J. Humphray, L. French, R. S. Evans, G. Bethel, A. Whittaker, J. L. Holden, O. T. Mccann, A. Dunham, C. Soderlund, C. E. Scott, D. R. Bentley, G. Schuler, H. C. Chen, W. H. Jang, E. D. Green, J. R. Idol, V. V. B. Maduro, K. T. Montgomery, E. Lee, A. Miller, S. Emerling, R. Kucherlapati, R. Gibbs, S. Scherer, J. H. Gorrell, E. Sodergren, K. Clerc-Blankenburg, P. Tabor, S. Naylor, D. Garcia, P. J. de Jong, J. J. Catanese, N. Nowak, K. Osoegawa, S. Z. Qin, L. Rowen, A. Madan, M. Dors, L. Hood, B. Trask, C. Friedman, H. Massa, V. G. Cheung, I. R. Kirsch, T. Reid, R. Yonescu, J. Weissenbach, T. Bruls, R. Heilig, E. Branscomb, A. Olsen, N. Doggett, J. F. Cheng, T. Hawkins, R. M. Myers, J. Shang, L. Ramirez, J. Schmutz, O. Velasquez, K. Dixon, N. E. Stone, D. R. Cox, D. Haussler, W. J. Kent, T. Furey, S. Rogic, S. Kennedy, S. Jones, A. Rosenthal, G. P. Wen, M. Schilhabel, G. Gloeckner, G. Nyakatura, R. Siebert, B. Schlegelberger, J. Korenburg, X. N. Chen, A. Fujiyama, M. Hattori, A. Toyoda, T. Yada, H. S. Park, Y. Sakaki, N. Shimizu, S. Asakawa, K. Kawasaki, T. Sasaki, A. Shintani, A. Shimizu, K. Shibuya, J. Kudoh, S. Minoshima, J. Ramser, P. Seranski, C. Hoff, A. Poustka, R. Reinhardt, and H. Lehrach. 2001. A physical map of the human genome. Nature 409:934-941.

Merks, J., D. Ducro-Steverink, and H. Feitsma. 2000. Management and genetic factors affecting fertility in sows. Reproduction in Domestic Animals 35:261-266.

Messer, L. A., L. Z. Wang, C. K. Tuggle, M. Yerle, P. Chardon, D. Pomp, J. E. Womack, W. Barendse, A. M. Crawford, D. R. Notter, and M. F. Rothschild. 1997. Mapping of the melatonin receptor 1a (MTNR1A) gene in pigs, sheep, and cattle. Mamm Genome 8:368-370.

Milan, D., J. P. Bidanel, P. Le Roy, N. Chevalet, N. Woloszyn, J. C. Caritez, H. Gruand, M. Bonneau, L. Lefaucheur, C. Renard, M. Vaiman, P. Mormede, C. Desautes, Y. Amigues, F. Bourgeois, J. Gellin, and L. Olliver. 1998. Current status of QTL detection in large white x meishan crosses in France. Proceedings of the 6th world congress on genetics applied to livestock production, Armidale NSW 2351, Australia, January 11-16, 1998.414-417.

Miller, R. 1997. Linkage mapping of plant and animal genomes. In: P. H. Dear (Ed.) Genome mapping - a practical approach. pp. 27-47. Oxford University Press.

Miller, R. D. and P. Y. Kwok. 2001. The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine. Hum Mol Genet 10:2195-2198.

Moller, M. J., R. Chaudhary, E. Hellmen, B. Hoyheim, B. P. Chowdhary, and L. Andersson. 1996. Pigs with the dominant white coat color phenotype carry a duplication of the KIT gene encoding the mast/stem cell growth factor receptor. Mamm Genome 7:822-830.

Montgomery, G. W., K. P. McNatty, and G. H. Davis. 1992. Physiology and molecular genetics of mutations that increase ovulation rate in sheep. Endocr Rev 13:309-328.

Montgomery, G. W., J. M. Penty, E. A. Lord, J. Brooks, and A. S. McNeilly. 1995a. The gonadotrophin-releasing hormone receptor maps to sheep chromosome 6 outside of the region of the FecB locus. Mamm Genome 6:436-438.

Montgomery, G. W., J. M. Penty, E. A. Lord, and M. F. Broom. 1995b. The search for the Booroola (FecB) mutation. J Reprod Fertil Suppl 49:113-121.

Moran, C. 1993. Microsatellite repeats in pig (Sus domestica) and chicken (Gallus domesticus) genomes. J Hered 84:274-280.

Moran, C. 1998. Molecular Genetics. In: M. F. Rothschild and A. Ruvinsky (Eds.) The Genetics of the Pig. pp. 141-143. CABI publishing.

Morton, N. E. 1991. Parameters of the Human Genome. Proc Natl Acad Sci U S A 88:7474-7476.

Mullins, D. E., F. W. Bazer, and R. M. Roberts. 1980. Secretion of a progesterone-induced inhibitor of plasminogen activator by the porcine uterus. Cell 20:865-872.

Newell, W., P. Sanseau, J. Riley, and A. Lyall. 1998. WEBMAP: radiation hybrid mapping on the WWW. Bioinformatics 14:825-826.

Nicholas, K. B. and H. B. Jr. Nicholas. GeneDoc: Analysis and Visualization of Genetic Variation. [version 2.5.006]. 1997.

Nomura, S., A. J. Wills, D. R. Edwards, J. K. Heath, and B. L. M. Hogan. 1988. Developmental expression of 2AR (osteopontin) and SPARC (osteonectin) RNA as revealed by in situ hybridisation. J Cell Biol 106:441-450.

O'Brien, S. J. and J. A. M. Graves. 1991. Report of the committee on comparative gene mapping. Cytogenet Cell Genet 58:1124-1151.

O'Brien, S. J., J. E. Womack, L. A. Lyons, K. J. Moore, N. A. Jenkins, and N. G. Copeland. 1993. Anchored reference loci for comparative genome mapping in mammals. Nat Genet 3:103-112.

Okamura, N., M. Tamba, H. J. Liao, S. Onoe, Y. Sugita, F. Dacheux, and J. L. Dacheux. 1995. Cloning of complementary-DNA encoding a 135-Kilodalton protein secreted from porcine corpus epididymis and its identification as an epididymis-specific alpha-mannosidase. Mol Reprod Dev 42:141-148.

Olliver, L. 1998. Genetic Improvement of the Pig. In: M. F. Rothschild and A. Ruvinsky (Eds.) The Gentics of the Pig. pp. 511-540.

Ollivier, L. and G. Bolet. 1981. Selection for Prolificacy in the Pig - Results of A 10 Generation Selection Experiment. Annales de Zootechnie 30:382.

Olson, M. V. 2001. The maps - Clone by clone by clone. Nature 409:816-818.

Olson, M. V. 2002. The human genome project: A player's perspective. J Mol Biol 319:931-942.

Olson, M. V. and A. Varki. 2003. Sequencing the chimpanzee genome: Insights into human evolution and disease. Nature Reviews Genetics 4:20-28.

Omigbodun, A., P. Ziolkiewicz, C. Tessler, J. R. Hoyer, and C. Coutifaris. 1997. Progesterone regulates osteopontin expression in human trophoblasts: A model of paracrine control in the placenta? Endocrinology 138:4308-4315.

Otsuka, F., Z. X. Yao, T. H. Lee, S. Yamamoto, G. F. Erickson, and S. Shimasaki. 2000. Bone morphogenetic protein-15 - Identification of target cells and biological functions. J Biol Chem 275:39523-39528.

Palmiter, R. D., R. L. Brinster, R. E. Hammer, M. E. Trumbauer, M. G. Rosenfeld, N. C. Birnberg, and R. M. Evans. 1982. Dramatic Growth of Mice That Develop from Eggs Micro-Injected with Metallothioneine-Growth Hormone Fusion Genes. Nature 300:611-615.

Pomp, D., S. Foster, M. A. Cushman, and G. Eisen. 1995. Identification of genes controlling quantitative reproduction traits in mice. Biol Reprod 52:61.

Pope, W. F., S. Xie, D. M. Broermann, and K. P. Nephew. 1990. Causes and consequences of early embryonic diversity in pigs. J Reprod Fertil Suppl 40:3-17.

Prince, C. W., T. Oosawa, W. T. Butler, M. Tomana, A. S. Bhown, M. Bhown, and R. E. Schrohenloher. 1987. Isolation, Characterization, and Biosynthesis of A Phosphorylated Glycoprotein from Rat Bone. J Biol Chem 262:2900-2907.

Proost, P., A. Wuyts, R. Conings, J. P. Lenaerts, A. Billiau, G. Opdenakker, and J. Van Damme. 1993. Human and bovine granulocyte chemotactic protein-2:

complete amino acid sequence and functional characterization as chemokines. Biochemistry 32:10170-10177.

Raeymaekers, P., K. Van Zand, L. Jun, M. Hoglund, J. J. Cassiman, H. Van den Berghe, and P. Marynen. 1995. A radiation hybrid map with 60 loci covering the entire short arm of chromosome 12. Genomics 29:170-178.

Rathje, T. A., G. A. Rohrer, and R. K. Johnson. 1997. Evidence for quantitative trait loci affecting ovulation rate in pigs. J Anim Sci 75:1486-1494.

Rettenberger, G., C. Klett, U. Zechner, J. Kunz, W. Vogel, and H. Hameister. 1995. Visualisation of the conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics 26:372-378.

Riquet, J., W. Coppieters, N. Cambisano, J. J. Arranz, P. Berzi, S. K. Davis, B. Grisart, F. Farnir, L. Karim, M. Mni, P. Simon, J. F. Taylor, P. Vanmanshoven, D. Wagenaar, J. E. Womack, and M. Georges. 1999. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. Proc Natl Acad Sci U S A 96:9252-9257.

Robert, C., M. F. Palin, N. Coulombe, C. Roberge, F. G. Silversides, B. F. Benkel, R. M. Mckay, and G. Pelletier. 1998. Backfat thickness in pigs is positively associated with leptin mRNA levels. Canadian Journal of Animal Science 78:473-482.

Roberts, R. M., S. Xie, and W. E. Trout. 1993. Embryo-Uterine Interactions in Pigs During Week-2 of Pregnancy. J Reprod Fertil171-186.

Robic, A., M. Dalens, N. Woloszyn, D. Milan, J. Riquet, and J. Gellin. 1994. Isolation of 28 new porcine microsatellites revealing polymorphism. Mamm Genome 5:580-583.

Robic, A., D. Milan, N. Woloszyn, J. Riquet, M. Yerle, M. Nagel, M. Bonnet, P. Pinton, M. Dalens, and J. Gellin. 1997. Contribution to the physically anchored linkage map of the pig. Anim Genet 28:94-102.

Rohrer, G. A. 1999. Mapping four genes from human chromosome 4 to porcine chromosome 8 further develops the comparative map for an economically important chromosome of the swine genome. Anim Genet 30:60-62.

Rohrer, G. A. 2000. Identification of quantitative trait loci affecting birth characters and accumulation of backfat and weight in a Meishan- White Composite resource population. J Anim Sci 78:2547-2553.

Rohrer, G. A., L. J. Alexander, Z. Hu, T. P. Smith, J. W. Keele, and C. W. Beattie. 1996. A comprehensive map of the porcine genome. Genome Res 6:371-391.

Rohrer, G. A., L. J. Alexander, J. W. Keele, T. P. Smith, and C. W. Beattie. 1994. A microsatellite linkage map of the porcine genome. Genetics 136:231-245.

Rohrer, G. A., J. E. Beever, M. F. Rothschild, L. Schook, R. Gibbs, and G. M. Weinstock. Porcine Genomic Sequencing Initiative. 2003. (see http://www.genome.iastate.edu/newsletter/PigWhitePaper.html)

Rohrer, G. A., J. J. Ford, T. H. Wise, J. L. Vallet, and R. K. Christenson. 1999. Identification of quantitative trait loci affecting female reproductive traits in a multigeneration Meishan-White composite swine population. J Anim Sci 77:1385-1391.

Rose, E. A. 1991. Applications of the polymerase chain reaction to genome analysis. FASEB J 5:46-54.

Rothberg, J. M., J. R. Jacobs, C. S. Goodman, and S. Artavanistsakonas. 1990. Slit - An extracellular protein necessary for development of midline glia and commissural axon pathways contains both EGF and LRR domains. Genes Dev 4:2169-2187.

Rothschild, M., C. Jacobson, D. Vaske, C. Tuggle, L. Wang, T. Short, G. Eckardt, S. Sasaki, A. Vincent, D. McLaren, O. Southwood, H. van der Steen, A. Mileham, and G. Plastow. 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proc Natl Acad Sci U S A 93:201-205.

Rothschild, M. F., L. A. Messer, and A. Vincent. 1997. Molecular approaches to improved pig fertility. J Reprod Fertil Suppl 52:227-236.

Rothschild, M. F., A. Vincent, C. Tuggle, G. A. Evans, T. Short, O. Southwood, R. Wales, and G. Plastow. 1998. A mutation in the prolactin receptor gene is associated with increased litter size in pigs. Anim Genet Suppl 29:69.

Rozen, S. and H. J. Skaletsky. 2000. Primer 3 on the WWW for general users and for biologist programmers. Meth Mol Biol 132:365-386.

Russel, A. J. F., S. A. Alexieva, and D. A. Elston. 1997. The effect of the introduction of the Thoka gene for fecundity on lamb production from Cheviot ewes. Anim Sci 64:503-507.

Saavedra, R. A., S. K. Kimbro, D. N. Stern, J. Schnuer, S. Ashkar, M. J. Glimcher, and C. I. Ljubetic. 1995. Gene-Expression and Phosphorylation of Mouse Osteopontin. Ann N Y Acad Sci 760:35-43.

Sabour, M. P., J. R. Chambers, A. A. Grunder, U. Kuhnlein, and J. S. Gavora. 1992. Endogenous Viral Gene Distribution in Populations of Meat-Type Chickens. Poultry Science 71:1259-1270.

Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. M. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S.

Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. S. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, and D. Altshuler. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928-933.

Safran, J. B., W. T. Butler, and M. C. Farach-Carson. 1998. Modulation of osteopontin post-translational state by 1,25- (OH)(2)-vitamin D-3 - Dependence on Ca2+ influx. J Biol Chem 273:29935-29941.

Sakurai, M., J. H. Zhou, M. Ohtaki, T. Itoh, Y. Murakami, and H. Yasue. 1996. Assignment of c-KIT gene to swine chromosome 8p12-p21 by fluorescence in situ hybridization. Mamm Genome 7:397.

Sayers, I., X. Chen, S. Ye, B. Beghe, and I. N. M. Day. 2000. MADGE and other single nucleotide polymorphism analysis methods: Application to the molecular genetic epidemiology of asthma and cardiovascular disease. In: A. Hajeer, J. Worthington, and S. John (Eds.) SNP and microsatellite genotyping: Markers for genetic analysis. pp. 13-34. Eaton Publishing.

Schiex, T., P. Chabrier, M. Bouchez, and D. Milan. Boosting EM for radiation hybrid and genetic mapping. Proceedings of the First Workshop on Algorithms in Bioinformatics, University of Aarhus, Denmark, August 28-31. 2001.

Schiex, T. and C. Gaspin. CarthaGene: Constructing and joining maximum likelihood genetic maps. Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology, Halkidiki, Greece, June., 258-268. 1997.

Schmitz, A., B. Chaput, P. Fouchet, M. N. Guilly, G. Frelat, and M. Vaiman. 1992. Swine Chromosomal DNA Quantification by Bivariate Flow Karyotyping and Karyotype Interpretation. Cytometry 13:703-710.

Seaton, G., C. S. Haley, S. A. Knott, M. Kearsey, and P. M. Visscher. 2002. QTL Express: mapping quantitative trait loci in of simple and complex pedigrees. Bioinformatics 18:339-340.

Senger, D. R., C. A. Perruzzi, and A. Papadopoulos. 1989. Elevated expression of secreted phosphoprotein-I (osteopontin, 2AR) as a consequence of neoplastic transformation. Anticancer Res 9:1291-1299.

Sham, P. 1998. The analysis of alleic associations. In: Statistics in human genetics. pp. 145-185. Arnold.

Short, T. H., M. F. Rothschild, O. I. Southwood, D. G. McLaren, A. deVries, H. vanderSteen, G. R. Eckardt, C. K. Tuggle, J. Helm, D. A. Vaske, A. J. Mileham, and G. S. Plastow. 1997. Effect of the estrogen receptor locus on reproduction and production traits in four commercial pig lines. J Anim Sci 75:3138-3142.

Simmen, R. C. M., F. A. Simmen, A. Hofig, S. J. Farmer, and F. W. Bazer. 1990. Hormonal-Regulation of Insulin-Like Growth-Factor Gene- Expression in Pig Uterus. Endocrinology 127:2166-2174.

Singer, D. S., L. J. Parent, and R. Ehrlich. 1987. Identification and DNA-sequence of an interspersed repetitive DNA element in the genome of the miniature swine. Nucleic Acids Res 15:2780.

Singh, K., M. W. Devouge, and B. B. Mukherjee. 1990. Physiological-Properties and Differential Glycosylation of Phosphorylated and Nonphosphorylated Forms of Osteopontin Secreted by Normal Rat-Kidney Cells. J Biol Chem 265:18696-18701.

Slonim, D., L. Kruglyak, L. Stein, and E. Lander. 1997. Building human genome maps with radiation hybrids. J Comput Biol 4:487-504.

Smith, J. and D. W. Burt. 1998. Parameters of the chicken genome (Gallus gallus). Anim Genet 29:290-294.

Smith, S. J., S. Cases, D. R. Jensen, H. C. Chen, E. Sande, B. Tow, D. A. Sanan, J. Raber, R. H. Eckel, and R. V. Farese. 2000. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. Nat Genet 25:87-90.

Smith, T. P. L., N. L. LopezCorrales, S. M. Kappes, and T. S. Sonstegard. 1997. Myostatin maps to the interval containing the bovine mh locus. Mamm Genome 8:742-744.

Sorensen, E. S., P. Hojrup, and T. E. Petersen. 1995. Posttranscriptional modifications of bovine osteopontin - Identification of 28 phosphorylation and 3 O-glycosylation sites. Protein Sci 4:2040-2049.

Sorensen, E. S. and T. E. Petersen. 1994. Identification of 2 Phosphorylation Motifs in Bovine Osteopontin. Biochem Bioph Res Co 198:200-205.

Soubrier, F., S. Martin, A. Alonso, S. Visvikis, L. Tiret, F. Matsuda, G. M. Lathrop, and M. Farrall. 2002. High-resolution genetic mapping of the ACE-linked QTL influencing circulating ACE activity. European Journal of Human Genetics 10:553-561.

Southern, E. M. 1975. Detection of specific sequences among DNA fragmants seperated by gel electrophoresis. J Mol Biol 98:503-517.

Souza, C. J. H., C. MacDougall, B. K. Campbell, A. S. McNeilly, and D. T. Baird. 2001. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. J Endocrinol 169:R1-R6.

Spearow, J. L., M. Barkley, C. L. Turner, O. M. Moeras, N. J. Alcaraz, T. Nguyen, A. Venkatesan, D. Huang, L. Sullivan, and B. C. Hugins. 2000. Higher-

resolution mapping and characterization of ovulation rate induced and aromatase activity induced QTL in mice. Biol Reprod 62:324.

Spearow, J. L., S. Keisner, L. Carr, M. Peters, J. Barthelow, J. Faridi, R. Freed, R. Bury, V. Patel, and M. Barkley. 1995. Use of molecular genetic markers to develop congenic strains and selected lines of mice with altered reproduction. Biol Reprod 52:61.

Spotter, A., H. Kuiper, C. Drogemuller, B. Brenig, T. Leeb, and O. Distl. 2002. Assignment of the porcine epidermal growth factor (EGF) gene to SSC8q2.3-q2.4 by fluorescence in situ hybridization and radiation hybrid mapping. Anim Genet 33:166-167.

Staden, R. 1996. The Staden sequence analysis package. Mol Biotechnol 5:233-241.

Stewart, E. A., K. B. McKusick, A. Aggarwal, E. Bajorek, S. Brady, A. Chu, N. Fang, D. Hadley, M. Harris, S. Hussain, R. Lee, A. Maratukulam, K. OConnor, S. Perkins, M. Piercy, F. Qin, T. Reif, C. Sanders, X. H. She, W. L. Sun, P. Tabar, S. Voyticky, S. Cowles, J. B. Fan, C. Mader, J. Quackenbush, R. M. Myers, and D. R. Cox. 1997. An STS-based radiation hybrid map of the human genome. Genome Res 7:422-433.

Tabas, J. A., M. Zasloff, J. J. Wasmuth, B. S. Emanuel, M. R. Altherr, J. D. McPherson, J. M. Wozney, and F. S. Kaplan. 1991. Bone morphogenetic protein: chromosomal localization of human genes for BMP1, BMP2A, and BMP3. Genomics 1991 Feb;9(2):283-9. 9:283-289.

Takagi, Y. and S. Sasaki. 1988. A Probable Common Disturbance in the Early Stage of Odontoblast Differentiation in Dentinogenesis-Imperfecta Type-I and Type-II. Journal of Oral Pathology & Medicine 17:208-212.

Takahashi, H., T. Awata, and H. Yasue. 1992. Characterization of Swine Short Interspersed Repetitive Sequences. Anim Genet 23:443-448.

The Arabidopsis Genome Inititive. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature 408:798-815.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the *nematode C-elegans*: A platform for investigating biology. Science 282:2012-2018.

Tumbleson, M. E. and L. B. Schook. 1996. Advances in swine in biomedical research. Plenum Press.

van der Steen, H., O. Southwood, A. Vries, T. Short, D. McLaren, M. Wei, and G. Plastow. 1997. Evidence of a new genetic marker for litter size in Meishan synthetic pigs. Proceedings of the International Conference on Animal Biotechnology, Beijing, China, June 11-14, 1997.17-20.

Vasseur, F., N. Helbecque, C. Dina, S. Lobbens, V. Delannoy, S. Gaget, P. Boutin, M. Vaxillaire, F. Lepretre, S. Dupont, K. Hara, K. Clement, B. Bihain, T.

Kadowaki, and P. Froguel. 2002. Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte- secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. Hum Mol Genet 11:2607-2614.

Vaughan, P. and T. V. McCarthy. 1998. A novel process for mutation detection using uracil DNA- glycosylase. Nucleic Acids Res 26:810-815.

Venter, J. C., M. D. Adams, *et al.* 2001. The sequence of the human genome. Science 291:1304-1351.

Vonnahme, K. A., J. R. Malayer, H. O. Spivey, S. P. Ford, A. Clutter, and R. D. Geisert. 1999. Detection of kallikrein gene expression and enzymatic activity in porcine endometrium during the estrous cycle and early pregnancy. Biol Reprod 61:1235-1241.

Vonnahme, K. A., M. E. Wilson, and S. P. Ford. 2002. Conceptus competition for uterine space: Different strategies exhibited by the Meishan and Yorkshire pig. J Anim Sci 80:1311-1316.

Vukasinovic, N., S. K. Denise, and A. E. Freeman. 1999. Association of growth hormone loci with milk yield traits in Holstein bulls. Journal of Dairy Science 82:788-794.

Waddington, D. 2000. Estimating the number of conserved segments between species using a chromosome based model. In: D. Sankoff and J. H. Nadeau (Eds.) Comparative genomics. pp. 321-332. Kluwer Academic Publishers.

Wagner, R. P. 1994. Understanding Inheritance: An Introduction to Classical and Molecular Genetics. In: N. G. Cooper (Ed.) The Human Genome Project: Deciphering the blueprint of heredity. pp. 1-64. University Science Books.

Walling, G. A., A. L. Archibald, J. A. Cattermole, A. C. Downing, H. A. Finlayson, D. Nicholson, P. M. Visscher, C. A. Walker, and C. S. Haley. 1998. Mapping of quantitative trait loci on porcine chromosome 4. Anim Genet 29:415-424.

Walling, G. A., S. C. Bishop, R. Pong-Wong, G. Gittus, A. J. F. Russel, and S. M. Rhind. 2002. Detection of a major gene for litter size in Thoka Cheviot sheep using Bayesian segregation analyses. Anim Sci 75:339-347.

Walling, G. A., P. M. Visscher, L. Andersson, M. F. Rothschild, L. Z. Wang, G. Moser, M. A. M. Groenen, J. P. Bidanel, S. Cepica, A. L. Archibald, H. Geldermann, D. J. De Koning, D. Milan, and C. S. Haley. 2000. Combined analyses of data from quantitative trait loci mapping studies: Chromosome 4 effects on porcine growth and fatness. Genetics 155:1369-1378.

Walter, M. A., D. J. Spillett, P. Thomas, J. Weissenbach, and P. N. Goodfellow. 1994. A method for constructing radiation hybrid maps of whole genomes. Nat Genet 7:22-28.

Warren, W., T. P. L. Smith, C. E. Rexroad, S. C. Fahrenkrug, T. Allison, C. L. Shu, J. Catanese, and P. J. de Jong. 2000. Construction and characterization of a new bovine bacterial artificial chromosome library with 10 genome-equivalent coverage. Mamm Genome 11:662-663.

Watanabe, T. K., M. T. Bihoreau, L. C. McCarthy, S. L. Kiguwa, H. Hishigaki, A. Tsuji, J. Browne, Y. Yamasaki, A. Mizoguchi-Miyakita, K. Oga, T. Ono, S. Okuno, N. Kanemoto, E. Takahashi, K. Tomita, H. Hayashi, M. Adachi, C. Webber, M. Davis, S. Kiel, C. Knights, A. Smith, R. Critcher, J. Miller, T. Thangarajah, P. J. R. Day, J. R. Hudson, Y. Irie, T. Takagi, Y. Nakamura, P. N. Goodfellow, G. M. Lathrop, A. Tanigami, and M. R. James. 1999. A radiation hybrid map of the rat genome containing 5,255 markers. Nat Genet 22:27-36.

Waterhouse, P., R. S. Parhar, X. J. Guo, P. K. Lala, and D. T. Denhardt. 1992. Regulated temporal and spacial expression of the calcium-binding proteins calcyclin and OPN (osteopontin) in mouse tissues during pregnancy. Mol Reprod Dev 32:315-323.

Waterston, R. H., K. Lindblad-Toh, E. *et al* 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-562.

Watson, J. D. and F. H. C. Crick. 1953. Molecular structure of nucleic acids: A structure for Deoxyribose Nucleic Acid. Nature 171:737-738.

Webel, S. K. and P. J. Dziuk. 1974. Effect of stage of gestation and uterine space on prenatal survival in the pig. J Anim Sci 38:960-963.

Wildt, D. E., S. Fujimoto, J. L. Spencer, and W. R. Dukelow. 1973. Direct ovarian observation in the pig by means of laparoscopy. J Reprod Fertil 35:541-543.

Wilkie, P. J., A. A. Paszek, C. W. Beattie, L. J. Alexander, M. B. Wheeler, and L. B. Schook. 1999. A genomic scan of porcine reproductive traits reveals possible quantitative trait loci (QTLs) for number of corpora lutea. Mamm Genome 10:573-578.

Wilson, M. E. Role of placental function in mediating conceptus growth and survival. Proceedings of the International Animal Agriculture and Food Science Conference, Indianapolis, IN, USA July 24 -28, 2001.

Wilson, M. E., N. J. Biensen, and S. P. Ford. 1999. Novel insight into the control of litter size in pigs, using placental efficiency as a selection tool. J Anim Sci 77:1654-1658.

Wilson, M. E., N. J. Biensen, C. R. Youngs, and S. P. Ford. 1998. Development of Meishan and Yorkshire littermate conceptuses in either a Meishan or Yorkshire uterine environment to day 90 of gestation and to term. Biol Reprod 58:905-910.

Wilson, M. E. and S. P. Ford. 2000. Effect of estradiol-17 beta administration during the time of conceptus elongation on placental size at term in Meishan pigs. J Anim Sci 78:1047-1052.

Wilson, T., X. Y. Wu, J. L. Juengel, I. K. Ross, J. M. Lumsden, E. A. Lord, K. G. Dodds, G. A. Walling, J. C. McEwan, A. R. O'Connell, K. P. McNatty, and G. W. Montgomery. 2001. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. Biol Reprod 64:1225-1235.

Womack, J. E., J. S. Johnson, E. K. Owens, C. E. Rexroad, J. Schlapfer, and Y. P. Yang. 1997. A whole-genome radiation hybrid panel for bovine gene mapping. Mamm Genome 8:854-856.

Womble, D. D. 2000. GCG: The Wisconsin package of sequence analysis programs. Meth Mol Biol 132:22.

Wrana, J. L., Q. Zhang, and J. Sodek. 1989. Full length cDNA sequence of porcine secreted phosphoprotein-1 (SPP-1, oseopontin). Nucleic Acids Res 17:10119.

Yamamoto, S., N. Hijiya, M. Setoguchi, K. Matsuura, T. Ishida, Y. Higuchi, and S. Akizuki. 1995. Structure of the osteopontin gene and its promoter. Ann N Y Acad Sci 760:44-58.

Yasue, H. and Y. Wada. 1996. A swine SINE (PRE-1 sequence) distribution in swine-related animal species and its phylogenetic analysis in swine genome. Anim Genet 27:95-98.

Yerle, M., P. Pinton, A. Robic, A. Alfonso, Y. Palvadeau, C. Delcros, R. Hawken, L. Alexander, C. W. Beattie, L. Schook, D. Milan, and J. Gellin. 1998. Construction of a whole-genome radiation hybrid panel for high- resolution gene mapping in pigs. Cytogenet Cell Genet 82:182-188.

Yoshitake, H., S. R. Rittling, D. T. Denhardt, and M. Noda. 1999. Osteopontin-deficient mice are resistant to ovariectomy-induced bone resorption. Proc Natl Acad Sci U S A 96:8156-8160.

Yu, J., S. N. Hu, J. *et al.* 2002. A draft sequence of the rice genome (Oryza sativa L. ssp indica). Science 296:79-92.

Zhang, Q., J. L. Wrana, and J. Sodek. 1992a. Characterization of the promoter region of the porcine opn (osteopontin, secreted phosphoprotein 1) gene. Identification of positive and negative regulatory elements and a 'silent' second promoter. Eur J Biochem 207:649-659.

Zhang, Z., M. Krause, and D. L. Davis. 1992b. Epidermal Growth-Factor Receptors in Porcine Endometrium - Binding Characteristics and the Regulation of Prostaglandin-e and Prostaglandin-F(2-Alpha) Production. Biol Reprod 46:932-936

Appendix I

Protocols for reagent preparation:

PCR and gel electrophoresis

- **20x TBE buffer**: 216g Tris Base, 110g Boric acid and 16.6g disodium EDTA per 1000ml. Use double distilled water and store in the dark.
- **$T_{10}E_1$**: 10ml 1M Tris and 2ml 0.5M EDTA per 1000ml (pH 8). Use double distilled water, autoclave and store at room temperature.
- **6x loading buffer**: Add a couple of grains of bromophenol blue to 25 ml of 40% sucrose
- **40% sucrose**: 10g sucrose per 25ml. Use distilled water, mix and heat to dissolve. Filter and autoclave.
- **50x EB buffer**: 242g Tris Base, 18.61g disodium EDTA and 57.1ml acetic acid per 1000ml. Use double distilled water.

Radiation hybrid mapping

- **1M Tris HCL pH 8.8**: (Molecular weight 121.14). 60.57g Tris Base per 500ml. Use distilled water, mix and adjust pH to 8.8 with concentrated hydrochloric acid. Autoclave.
- **0.5M EDTA pH 8.0**: (Molecular weight 372.24). 37.224g EDTA (disodium salt) per 200ml. Use distilled water, mix and titrate to pH 8.0 with sodium hydroxide pellets. Autoclave.
- **$T_{10}E_{0.1}$** : 10ml 1M Tris and 200µl 0.5M EDTA per 1000ml (pH 8). Use double distilled water, autoclave and store at room temperature.
- **0.845 mg/ml Creosol red in $T_{10}E_{0.1}$**: 0.0845g creosol red per 100ml $T_{10}E_{0.1}$. Autoclave and store in dark at 4° C
- **1 M sodium hydroxide (NaOH):** (Molecular weight 40). 4g NaOH per 100ml. Use distilled water and stir to dissolve.
- **Dilution buffer:** 320ml $T_{10}E_{0.1}$, 5.2ml 0.845 mg/ml Creosol red in $T_{10}E_{0.1}$ and 0.56ml 1M NaOH per 1000ml. Use distilled water and autoclave.

Southern Hybridisation

- **Aqueous Prehybridisation/Hybridisation (APH) solution**: 7% w/v SDS, 0.5M Na_2HPO_4, pH 7.2 and 1mM EDTA, stored at 65° C
- **1M disodium hydrogen orthophosphate (Na_2PO_4) (pH 7.2)**:142g anhydrous Na_2HPO_4 and 8ml 85% orthophosphoric acid. Use double distilled water, filter, autoclave and store at room temperature
- **Wash 1**: 2x SSC, pH 7.0 and 0.1% SDS
- **Wash 2**: 0.5x SSC, pH 7.0 and 0.1% SDS
- **20x SSC (pH 7.0)** :175.3g NaCl and 88.2g NaCitrate per 1000ml. Use double distilled water, autoclave and store at room temperature

Luria-Bertani media

- **LB bottom agar**: 10g tryptone, 5g yeast extract, 5g NaCl and 15g bacto agar per 1000ml. Autoclave and store at 60° C or re-melt when required.
- **LB broth**: 10g tryptone, 5g yeast extract and 5g NaCl per 1000ml. Autoclave

Microarray

- **4x salt buffer**: 195.6 ml Na_2HPO_4, 4.4 ml NaH_2PO_4 and 80 µl 10 % SDS
- **20x SSC (pH 7.0)** :175.3g NaCl and 88.2g NaCitrate per 1000ml. Use double distilled water, autoclave and store at room temperature

Appendix II

Alignment of complete sequence of *SPP1* from Large White (LW) and Meishan (MS) breed origin. Key features of the gene and differences between the two sequences are highlighted.

```
            *       20          *       40          *       60          *       80          *       100
LW : GAATTCACTCGTCTTTCCTTTGAGGGAGACCAGCTCTTGAGCGAGTGTGGGAAGCGGGGAAGGAGCCCATCACGTCCACCTGCGGTTGCTAAAGACAACA :   100
MS : ----------GTCTTTCCTTTGAGGGAGACCAGCTCTTGAGCGAGTGTGGGAAGCGGGGAAGGAGCCCATCACGTCCACCTGCGGTTGCTAAAGACAACA :    90

            *       120         *       140         *       160         *       180         *       200
LW : GAGCAGAAAAGAACGCTCTGCTTCTCTTGGCCTCCGTGTTCCCTGTTAATGTGTAGCGCGTCGTTGTTGGGAAATAGTTCCTCACCTGACTTTCCAAGAA :   200
MS : GAGCAGAAAAGAACGCTCTGCTTCTCTTGGCCTCCGTGTTCCCTGTTAATGTGTAGCGCGTCGTTGTTGGGAAATAGTTCCTCACCTGACTTTCCAAGAA :   190

            *       220         *       240         *       260         *       280         *       300
LW : ATGGAGGGCCTCACAGTTGTTTGATGGCTCGGTCATTAAATGCATGATCGTTCCGTCCTGCCGGAGTCACTGACGGAACCAGACCGTGGTCTCAGGTCCT :   300
MS : ATGGAGGGCCTCACAGTTGTTTGATGGCTCGGTCATTAAATGCATGATCGTTCCGTCCTGCCGGAGTCACTGACGGAACCAGACCGTGGTCTCAGGTCCT :   290

            *       320         *       340         *       360         *       380         *       400
LW : TCTCCGAAATGCTGCCATCGTGTGGCACCTCGGAGCCATGACCGGAAGAGCCCTATGGGTCATATGGTTCAGCGCAGGGTGGCTGGACTCCAGCAGAATC :   400
MS : TCTCCGAAATGCTGCCATCGTGTGGCATCTCGGAGCCATGACCGGAAGAGCCCTATGGGTCATATGGTTCAGCGCAGGGTGGCTGGGCTCCAGCAGAATC :   390

            *       420         *       440         *       460         *       480         *       500
LW : TATTCCTATAACTGTCTACGTTCATATTAGACCATTCCCGTGGGCACAGAGTAAACCATAGTGAATCCTGCGGAAATCTTGGCTGTTTTTAGAATTTGTG :   500
MS : TATTCCTATAACTGTCTACGTTCATATTAGACCATTCCCGTGGGCACAGAGTAAACCATAGTGAATCCTGCGGAAATCTTGGCTGTTTTTAGAATTTGTG :   490

            *       520         *       540         *       560         *       580         *       600
LW : AACTTCCCTCCACGACACTGACAATATGACAAACTTAAGTGAGTAATGTGTTTAAAGGGAAAAAACAGTTTTTAAGAGCAGAAATGAGAAATCTGGTTTT :   600
MS : AACTTCCCTCCACGACACTGACAATATGACAAACTTAAGTGAGTAATGTGTTTAAAGGGAAAAAACAGTTTTTAAGAGCAGAAATGAGAAATCTGGTTTT :   590

            *       620         *       640         *       660         *       680         *       700
LW : GCAACCTGATAACCTGTGTACTTTATATTTTATAGAGACAGCTGCCATCCGAATTATTAAAATGTGTGCTTAGCACTCACAAAGCAACAATATCAATTCA :   700
MS : GCAACCTGATAACCTGTGTACTTTATATTTTATAGAGACAGCTGCCATCCGAATTATTAAAATGTGTGCTTAGCACTCACAAAGCAACAATATCAATTCA :   690

            *       720         *       740         *       760         *       780         *       800
LW : TTTGAAATGCATCCAATTTGAAAGTGACATTTCCACCCAACAAATCTATGGATTAATTATACTAGCACTGCAAAAAAATTGCTTAACCTGTATTTATGTT :   800
MS : TTTGAAATGCATCCAATTTGAAAGTGACATTTCCACCCAACAAATCTATGGATTAATTATACTAGCACTGCAAAAAAATTGCTTAACCTGTATTTATGTT :   790
```

```
                *       820         *       840         *       860         *       880         *       900
LW  : AAAATGAATATTTGGTAAATAGGAACTGACTCCTTAGGACTAATAATAAATAGGACCATTTATCTTCAGTCTCATCTTACACGTGAGTCTTACAGTGGAG  :   900
MS  : AAAATGAATATTTGGTAAATAGGAACTGACTCCTTAGGACTAATAATAAATAGGACCATTTATCTTCAGTCTCATCTTACACGTGAGTCTTACAGTGGAG  :   890

                *       920         *       940         *       960         *       980         *      1000
LW  : GTGTGAGATAAATGACTACTGCAAGCTCCTTTCACAACTGAGAAAGGGAGATGAAGAGGGTAAGTAACGTCAAACAATATTAAATGTTTCAAATGGGCTC  :  1000
MS  : GTGTGAGATAAATGACTACTGCAAGCTCCTTTCACAACTGAGAAAGGGAGATGAAGAGGGTAAGTAACGTCAAACAATATTAAATGTTTCAAATGGGCTC  :   990

                *      1020         *      1040         *      1060         *      1080         *      1100
LW  : AGAGCTCTACTACCCTGAACTTTGTTCCAATATTCAACTTTCATCTCCAGTTTTCTTTCAAACACTTTTTCAATACCCAGTAAAGTTTTTTAATATAAAA  :  1100
MS  : AGAGCTCTACTACCCTGAACTTTGTTCCAATATTCAACTTTCATCTCCAGTTTTCTTTCAAACACTTTTTCAATACCCAGTAAAGTTTTTTAATATAAAA  :  1090

                *      1120         *      1140         *      1160         *      1180         *      1200
LW  : TTTTATATTTAATTTTCATTTAAGTAACCAACTTTATATATCCTGGGAAAAAACACTAGAAAAAGACAGTTCAGAAACCTAATCCATTCCCGCAGATGTG  :  1200
MS  : TTTTATATTTAATTTTCATTTAAGTAACCAACTTTATATATCCTGGGAAAAAACACTAGAAAAAGACAGTTCAGAAACCTAATCCATTCCCGCAGATGTG  :  1190

                *      1220         *      1240         *      1260         *      1280         *      1300
LW  : TGCCAATTAGCCTGTTGATGTGCACAGTTTAAAAAGGCTACATCTGGAGTTCCCATTTGTGGCTCAGCGATAATGCATCTGACTACTATCCATGAGGACA  :  1300
MS  : TGCCAATTAGCCTGTTGATGTGCACAGTTTAAAAAAGCTACATCTGGAGTTCCCATTTGTGGCTCAGCGATAATGCATCTGACTACTATCCATGAGGACA  :  1290

                *      1320         *      1340         *      1360         *      1380         *      1400
LW  : CAGGTTCGATCTCTGGCCTCCATCAGTGGGTTAAGGATTCAGCATTGCTATGACCTATGGTGTAGTTCGCAGACATGGCTCCAATCTGGCGTGGCTGTGG  :  1400
MS  : CAGGTTCGATCTCTGGCCTCCATCAGTGGGTTAAGGATTCAGCATTGCTATGACCTTTGGTGTAGTTCGCAGACATGGCTCCAATCTGGCGTGGCTGTGG  :  1390

                *      1420         *      1440         *      1460         *      1480         *      1500
LW  : CTGTGGTGCAGGCCAGCAGGTGCAGCTCCGATTCAACCCCCTAGCCTGGGAACTTCCATGTGAAGCATGTGCAGCCCTTAAAAAAAAAAACAACCAAAAAAC  :  1500
MS  : CTGTGGTGCAGGCCAGCAGGTGCAGCTCCGATTCAGCCCCCTAGCCTGGGAACTTCCATGTGAAACATGTGCAGCCCTTAAAAAAAAAAACAACCAAAAAAC  :  1490

                *      1520         *      1540         *      1560         *      1580         *      1600
LW  : CTTACATCTGGTCGCATTTTAATAGCCTATTCCATTTAAAGATTCACAAATCATGACTACCTGTTCCTCCTAAAAATTTTAATAATAATTAACACATTAT  :  1600
MS  : CTTACATCTGGTCGCATTTTAATAGCCTATTCCATTTAAAGATTCACAAATCATGACTACCTGTTCCTCCTAAAAATTTTAATAATAATTAACACATTAT  :  1590

                *      1620         *      1640         *      1660         *      1680         *      1700
LW  : ACAGTTTAAAATATGCAAGGCATTTGCATATATGTGCCAAGTATCATTAGTTTTACACACACTAATTCACTTAAACCTCTCAAAACCTCCATGAACAAGG  :  1700
MS  : ACAGTTTAAAATATGCAAGGCATTTGCATATATGTGCCAACTATCATTAGTTTTACACACACTAATTCACTTAAACCTCTCAAAACCTCCATGAACAAGG  :  1690
```


1720
1740
1760
1780
1800
LW : TAATAGCATGCATATTTGACAGATGAAGAAACTGGTATAGAAAGGCTCAGTGATTTGCCCAAAGTTACAGGGCTAATTAGCAGCAGGATGGCAGGCAAAG : 1800
MS : TAATAGCATGCATATTTGACAGATGAAGAAACTGGTATAGAAAGGCTCAGTGATTTGCCCAAGGTTACAGGGCTAATTAGCAGCAGGATGGCAGGCAAAG : 1790
1820
1840
1860
1880
1900
LW : CCACGTCTTGTTTCTGAAGCATCACTTTCAAGCAGTAACCCTGCCCTCAGTCAGAAACTGCTTGACTTCTGCCACATTGAGAATAAATTACCATTCTTCT : 1900
MS : CCACGTCTTGTTTCTGAAGCATCACTTTCAAGCAGTAACCCTGCCCTCAGTCAGAAACTGCTTGACTTCTGCCACATTGAGAATAAATTACCATTCTTCT : 1890
1920
1940
1960
1980
2000
LW : GTCTGACATAGCAATTATCAGATTCATCACCTCCAAACCATTCAAAAATCCATGTCCTTGAAAGGTCTCGGATTTACAATGGTAAATCCGATTTAATAAA : 2000
MS : GTCTGACATAGCAATTATCAGATTCATCACCTCCAAACCATTCAAAAATCCATGTCCTTGAAAGGTCTCGGATTTACAATGGTAAATCCGATTTAATAAA : 1990
2020
2040
2060
2080
2100
LW : CATGAACGTGGATTTGAACGTGACCAATGCCACTTGTCTAAATGACAAGCTGCTGCACATTTGGAATCACAAAACAGAGCTTTAGTCGAAATGGAGAGAA : 2100
MS : CATGAACGTGGATTTGAACGTGACCAATGCCACTTGTCTAAATGACAAGCTGCTGCACATTTGGAATCACAAAACAGAGCTTTAGTCGAAATGGAGAGAA : 2090
2120
2140
2160
2180
2200
LW : GCTGGCAGAACTCTTTGTGTGCCTGAACAATATAGCCTTGTCGCCAGAAATAAGTACAACTGTAAGGCAAACTTGGTTGCATATATTACCTTAATCTGTT : 2200
MS : GCTGGCAGAACTCTTTGTGTGCCTGAACAATATAGCCTTGTCGCCAGAAATAAGTACAACTGTAAGGCAAACTTGGTTGCATATATTACCTTAATCTGTT : 2190
2220
2240
2260
2280
2300
LW : ATTTCAACCAAAAGAAACAAAAATCCATCCTATTTAATTCTATTTAGGACTGATTTCTGTAATATTTTTTCTTCTAATTCAGAGTTAATATAGTCTTTGT : 2300
MS : ATTTCAACCAAAAGAAACAAAAATCCATCCTATTTAATTCTATTTAGGACTGATTTCTGTAATATTTTTTCTTCTAATTCAGAGCTAATATAGTCTTTGT : 2290
2320
2340
2360
2380
2400
LW : AAGAACTAGTCATTTCTCTCCCCGCCCCCCGCTTTTTTAGGGGACCCAGAGATGCCCTTCCAGGATGCTGAGAGCCAATCCTATTCACGAAAAAGCTAGT : 2400
MS : AAGAACTAGTCATTTCTCTCCCCGCCCCCCGCTTTTTTAGGGGACCCAGAGATGCCCTTCCAGGATGCTGAGAGCCAATCCTATTCACGAAAAAGCTAGT : 2390
2420
2440
2460
2480
2500
LW : TAATGACATTGTACATAAGTAATGTTTTAACTGCAGAGGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT---------- : 2485
MS : TAATGACATTGTACATAAGTAATGTTTTAACTGCAGAGGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT : 2490
Microsatellite repeat
2520
2540
2560
2580
2600
LW : ----TGTAACCACAAAACCAGAGGGGGAAGTGTGGGAGCAAGTGGGTTGGGCAGTGGCACAAACCTCATGACACATCTCTCCGCCCCCTGTGTTGGTGG : 2580
MS : TGTGTTGTAACCACAAAACCAGAGGGGGAAGTGTGGGAGCAAGTGGGTTGGGCAGTGGCACAAACCTCATGACACATCTCTCCGCCCCCTGTGTTGGTGG : 2590


* 2620 * 2640 * 2660 * 2680 * 2700
LW : CGGATGTCTGGTGCAGCCTTTAAATTCAGGGGAGGTCCAGGCTGTCAGCAGCGAGCGGAGGCCAGAGGGCAGCACTGACAGCCGCATCAGCATTGCTCCC : 2680
MS : CGGATGTCTGGTGCAGCCTTTAAATTCAGGGGAGGTCCAGGCTGTCAGCAGCGAGCGGAGGCCAGAGGGCAGCACTGACAGCCGCATCAGCATTGCTCCC : 2690
EXON 1
* 2720 * 2740 * 2760 * 2780 * 2800
LW : GGGACTGGACTCTTCGCGGGGCTGCAGACCAAGGTAAGCCTACAGCCCGCTGATGGTTGCTGTCGTAGGTAGGCACTGTTTCATTTCACGGGGAAGGTCA : 2780
MS : GGGACTGGACTCTTCGCGGGGCTGCAGACCAAGGTAAGCCTACAGCCCGCTGATGGTTGCTGTCGTAGGCAGGCACTGTTTCATTTCACGGGGAAGGTCA : 2790
* 2820 * 2840 * 2860 * 2880 * 2900
LW : AGTTGTCAGGAGAAGGGTTTAACCGCAAACGCTGGTAATCAGACTTAAAATTGTTCTAAGAGTCTTTCCAACAAGCATGTCAGATCCAGGGCAGATCTCC : 2880
MS : AGTTGTCAGGAGAAGGGTTTAACCGCAAACGCTGGTAATCAGACTTAAAATTGTTCTAAGAGTCTTTCCGACAAGCATGTCAGATCCAGGGCAGATCTCC : 2890
* 2920 * 2940 * 2960 * 2980 * 3000
LW : TGCGGGAAAGGTGACTGGCTATTTTGAAAGACAGTCAAATATAAAACTTAAAAATATGTCCACGGAGTCCTCAAAAGAATTATAACTACTTCTTGCCATC : 2980
MS : TGCGGGAAAGGTGACTGGCTATTTTGAAAGACAGTCAAATATAAAACTTAAAAATATGTCCACGGAGTCCTCAAAAGAATTATAACTACTTCTTGCCATC : 2990
* 3020 * 3040 * 3060 * 3080 * 3100
LW : AGAAAAAAAATAAAATTCTATGCCTATGTATTATATGTAGCTACGATGGTATCTGCATCATTTTAAATGGACAGTTGATGGTGGAGACATTTAGAAGGAA : 3080
MS : AGAAAAAAAATAAAATTCTATGCCTATGTATTATATGTAGCTACGATGGTATCTGCATCATTTTAAATGGACAGTTGATGGTGGAGACATTTAGAAGGAA : 3090
* 3120 * 3140 * 3160 * 3180 * 3200
LW : AATAATATTTAGTGGTAATGTTAAGCTGGTGATAATGCTAATTTTTAATGACTTGACTTACCTGTTTTCAAGGGAAAGAAACCCTTTCTGAATATTTTCA : 3180
MS : AATAATATTTAGTGGTAATGTTAAGCTGGTGATAATGCTAATTTTTAATGACTTGACTTACCTGTTTTCAAGGGAAAGAAACCCTTTCTGAATATTTTCA : 3190
* 3220 * 3240 * 3260 * 3280 * 3300
LW : CCCCTGTATTTAGCTGTTAACATTTTCACCAAAATACCCACATGACATTATGAAGACTTACAAATAGAAAGGCTGTGAACTGCTCTCAGTGTTTAATTTT : 3280
MS : CCCCTGTATTTAGCTGTTAACATTTTCACCAAAATACCCACATGACATTATGAAGACTTACAAATAGAAAGGCTGTGAACTGCTCTCAGTGTTTAATTTT : 3290
* 3320 * 3340 * 3360 * 3380 * 3400
LW : TCATTTCAAATTTTAGAACACCTTACTTAAATTACTAACTCTAGAGACAGCTTCATTTCACTTAAGTAGCACCTTTTAAATAATTTAAGCTGAAAAATCG : 3380
MS : TCATTTCAAATTTTAGAACACCTTACTTAAATTACTAACTCTAGAGACAGCTTCATTTCACTTAAGTAGCACCTTTTAAATAATTTAAGCTGAAAAATCG : 3390
* 3420 * 3440 * 3460 * 3480 * 3500
LW : CCCTTGAAATGCATGCTGGAAAATGGAGACAGCAAGTTTCTTTCTCTCTTTCCTTTTATTTTCCCTCTTTCTCTTTGTATTTTTCGTCTTTGAAAAAAAA : 3480
MS : CCCTTGAAATGCATGCTGGAAAATGGAGACAGCAAGTTTCTTTCTCTCTTTCCTTTTATTTTCCCTCTTTCTCTTTGTATTTTTCGTCTTTGAAAAAAAA : 3490


LW : TGTGTTCCCCTTCTAGCTTATTATTTTAATTAAATTACTGTTGATCTGTTTTTAGGTTTAGATGGCTGGAGATATCGGGTAGTGATGGCATATCTCTGA : 3579
MS : ATGTGTTCCCCTTCTAGCTTATTATTTTAATTAAATTACTGTTGATCTGTTTTTAGGTTTAGATGGCTGGAGATATCGGGTAGTGATGGCATATCTCTGA : 3590
LW : AACTCTACATTTTAAAGGGGACTAAATAAGACTTGTATGTAATCCCCCTCTCTCTTGCCTAACAGTAAGAGATGGAAAATAGAGGTGCCCTAACAAATAT : 3679
MS : AACTCTACATTTTAAAGGGGACTAAATAAGACTTGTATGTAATCCCCCTCTCTCTTGCCTAACAGTAAGAGATGGAAAATAGAGGTGCCCTAACAAATAT : 3690
LW : TAACTCAAAGGATCATAAAATTAAAAAGAAAAAATTTTTCTCTAAGTAGTAGAGAGTATTTCTATAGGAAAAATATATATATATATTTTTTCGTGATTA : 3779
MS : TAACTCAAAGGATCATAAAATTAAAAAGAAAAAATTTTTCTCTAAGTAGTAGAGAGTATTTCTATAGGAAAAATATATATATATATTTTTTCGTGATTA : 3790
LW : TTTTGTAATGTGGTGGCTTGAAAAGATGTCATTGTTTTAACCTAGGAGAAGATCAAATATTTCTTACAAAATATTTTGCAGGAAAATCATTACCATCAGA : 3879
MS : TTTTGTAATGTGGTGGCTTGAAAAGATGTCATTGTTTTAACCTAGGAGAAGATCAAATATTTCTTACAAAATATTTTGCAGGAAAATCATTACCATCAGA : 3890
EXON 2
Start
LW : ATTGCAGTGATAGCCTTCTGCCTCTGGGGCTTCGCCTCTGCCCTTCCAGTGAGTACAGCTGAATCTTAAACAGAATTCCCCCAAATAAATGAATTGTGTG : 3979
MS : ATTGCAGTGATAGCCTTCTGCCTCTGGGGCTTCGCCTCTGCCCTTCCAGTGAGTACAGCTGAATCTTAAACAGAATTCCCCCAAATAAATGAATTGTGTG : 3990
Start secreted protein
LW : CTTCAGTGTGCCAGGAACTCATTCTCACTGTAATCTTTCTTCCCATGTTCTATTTCAAAGGTTAAACAGACTAATTCTGGCAGCTCGGAGGAAAAGCTGG : 4079
MS : CTTCAGTGTGCCAGGAACTCATTCTCACTGTAATCTTTCTTCCCATGTTCTATTTCAAAGGTTAAACAGACTAATTCTGGCAGCTCGGAGGAAAAGCTGG : 4090
EXON 3
LW : TGAGTATCTTTAAGGTTTATCATCTGGTTAAATTGTCGATGCCACTGGTGCAAGATATACTACAAACATACGGCAATTTTACTTATTTCATAGTTAAACA : 4179
MS : TGAGTATCTTTAAGGTTTATCATCTGGTTAAATTGTCGATGCCACTGGTGCAAGATATACTACAAACATATGGCAATTTTACTTATTTCATAGTTAAACA : 4190
LW : AAATAAGTTAGAAGTATGTATTGAATGCCTGCTACGGTTCTAGGCCAGTACCACGCACAGGATTGCTTTTAATTACTGTCATCTCCTAAAACCAAACTAA : 4279
MS : AAATAAGTTAGAAGTATGTATTGAATGCCTGCTACGGTTCTAGGCCAGTACCACGCACAGGATTGCTTTTAATTACTGTCATCTCCTAAAACCAAACTAA : 4290
LW : ATATTTTAATAGAATAAAATCTTGCTCTGGGCCATTATCATTTACGTAAAATGAGTTCTATGAACTTGTGTGGTTTTTTGTTCCCATGAGTCAACACACT : 4379
MS : ATATTTTAATAGAATAAAATCTTGCTCTGGGCCATTATCATTTACGTAAAATGAGTTCTATGAACTTGTGTGGTTTTTTGTTCCCATGAGTCAACACACT : 4390

```
                *      4420          *      4440          *      4460          *      4480          *      4500
LW : GTTCCAGTAGCACGAAATTCCATCTTTTGCTGTACTTGGGGGCAGGTCAAGTTTAAAATTCACCTCACACATTGGTGAGTTGGTATAACCAGAACAGTCC :  4479
MS : GTTCCAGTAGCACGAAATTCCATCTTTTGCTGTACTTGGGGGCAGGTCAAGTTTAAAATTCACCTCACATATTGGTGGGTTGGTATAACCAGAACAGTCC :  4490

                *      4520          *      4540          *      4560          *      4580          *      4600
LW : CAAGAGCTCCTAAGATATTTGCTGAGAAGGCCATCATAAAAAGAAAATCTATAGGCAGCAAGGTCCTACTGCAGAGCAAAGGGAACTATATCCAATCTCC :  4579
MS : CAAGAGCTCCTAAGATATTTGCTGAGAAGGCCATCATAAAAAGAAAATCTATAGGCAGCAAGGTCCTACTGCAGAGCAAAGGGAACTATATCCAATCTCC :  4590

                *      4620          *      4640          *      4660          *      4680          *      4700
LW : TGGGATAGACCAAGATGGAAAATCATATTTTAAAAGAATTTACTGAGTCCATTTGCAGTACAGCAGAAATTGGCACAACATCAATAAACCAACTATACATT :  4678
MS : TGGGATAGACCACGATGGAAAATAATATTTTAAAAGAATTTACTGAGTCCGTTTGCAGTACAGCAGAAATTGGCACAACATCGTAAACCAACTATACATT :  4690

                *      4720          *      4740          *      4760          *      4780          *      4800
LW : AATTTTTTAAAGAGAACAAAAGGAATGATATAAAATTTCTATTGATTATAGGGAAAAAAGGAAGTCTATTTATGCTGTTTATATCCGCTGTGCTTGTCAG :  4778
MS : AATTTTTTAAAGAGAACAAAAGGAATGATATAAAATTTCTATTGATTATAGGGCAAAAAGGAAGTCTATTTATGCTGTTTATATCCGCTGTGCTTGTCAG :  4790

                *      4820          *      4840          *      4860          *      4880          *      4900
LW : ATGACTTCAGATCACATTTTTCTCACCAGGCATAAAACCAAGGCAAACTCTTGATTACAATGATTTGTAGCGCCATTGATGTACTGCCCCTCCCACCAGA :  4878
MS : ATGACTTCAGATCACATTTTTCTCACCAGGCGTAAAACCAAGGCGAACTCTTGATTACAATGATTTGTAGCGCCATTGATGTACTGCCCCTCCCACCAGA :  4890

                *      4920          *      4940          *      4960          *      4980          *      5000
LW : GGGGAAACCGTGCCAGCCAGGCAACAGGCGGGCATTGTCCCAGGGAGCTTGGACAAAAAGGCACACAGAGTTCAATTCCAGAAGAACAGAATAAAGGCCA :  4978
MS : GGGGAAACCGTGCCAGCCAGGCAACAGGCGGGCATTGTCCCAGGGAGCTTGGACAAAAAGGCACACAGAGTTCAATTCCAGAAGAACAGAATAAAGGCCA :  4990

                *      5020          *      5040          *      5060          *      5080          *      5100
LW : GTAGAGAGCTGCCTGGGGTCACACCGCAGTGAGATGGCTTAAAGAAGATGTTAGGAGACGTATCGCTGAGTGTGTTTGGGCGTGCAAGTGGGTGTGTGTG :  5078
MS : GTAGAGAGCTGCCTGGGGTCACACCGCAGTGAGATGGCTTAAAGAAGATGTTAGGAGACGTATCGCTGCGTGTGTTTGGGCGTGCAAGTGGGTGTGTGTG :  5090

                *      5120          *      5140          *      5160          *      5180          *      5200
LW : AGTGTGTCTGTTTTCCAAACTGAGTTAAGAGCACAAGTACTGAATCACAGTAAGAAGCCAGAAAAGTAATATTTGCAGAGAAGACAGGGGTCACGCTCTT :  5178
MS : AGTGTGTCTGTTTTCCAAACTGAGTTAAGAGCACAAGTACTGAATCACAGTAAGAAGCCAGAAAAGTAATATTTGCAGAGAAGACAGGGGTCAAGCTCTT :  5190

                *      5220          *      5240          *      5260          *      5280          *      5300
LW : TAAGGCAGGACTTTGTGCCTTTTTAATAGACTAAACTATGGCCCAGGAAGCCGTGGCAGTGTCAGTTGTTCACTGACACTTGGGAGATAAGTAAAGAGTT :  5278
MS : TAAGGCAGGACTTTGTGCCTTTTTAATAGACTAAACTATGGCCCAGGAAGCCGTGGCAGTGTCAGTTGTTCACTGACACTTGGGAGATAAGTAAAGAGTT :  5290
```

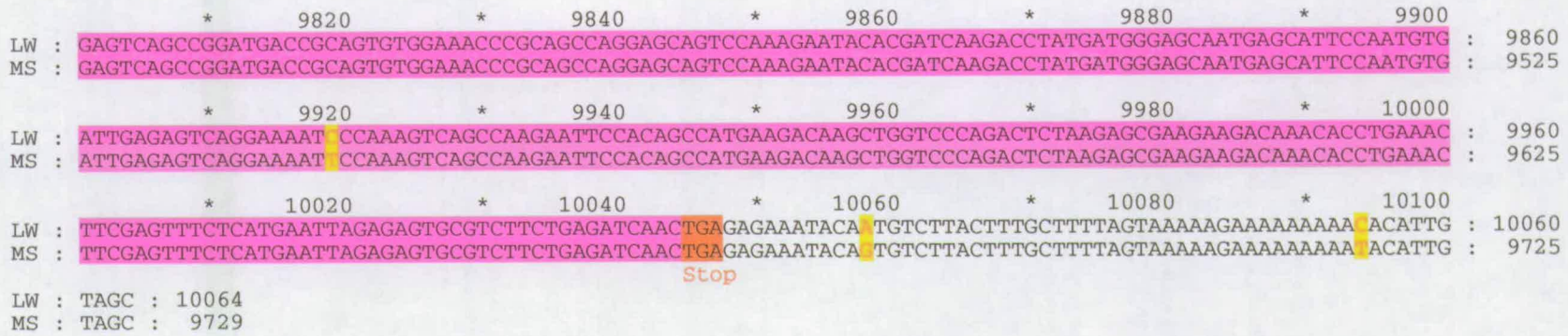

* 5320 * 5340 * 5360 * 5380 * 5400
LW : GAGCCTTCTAGAATCCTAGAAATCTAGAGCCGGTAGGAAATATGGAAAGTCTCTAAAGAGAACCTCCAAAGAGGAACACTGGTTACCTATGAGTGATTTT : 5378
MS : GAGCCTTCTAGAATCCTAGAACTC-------GGTAGGAAATATGGAAAGTCTCTAAAGAGAACCTTCAAAGAGGAACACTGGTTACCTATGAGTGATTTT : 5383
* 5420 * 5440 * 5460 * 5480 * 5500
LW : AGTTTGTTTTTCTTATCCATATATTCTTTCTTTCTGTCATTTTAGGGCTGCACCAGCAGCATAGGGAGGCTCCCAAGCTAGGGGTCGAATCAGAGCTGCA : 5478
MS : AGTTTGTTTTTCTTATCCATATCTTCTTTCTTTCTGTCATTTTAGGGCTGCACCAGCAGCATAGGGAGGCTCCCAAGCTAGGGGTCGAATCAGAGCTGCA : 5483
* 5520 * 5540 * 5560 * 5580 * 5600
LW : GCAGCTGACCTACACCACAGCCACACAATGCCAGATCCAAGTTGTATGTGCAGTCTCCACTGCAGCTCATGGCAATGCCGGATCCTTGACCCACTGAGTG : 5578
MS : GCAGCTGACCTACACCACAGCCACACAA----------GTTGTATGTGCAGTCTCCACTGCAGCTCACGGCAATGCCGGATCCTTGACCCACTGAGTG : 5571
* 5620 * 5640 * 5660 * 5680 * 5700
LW : AGGCCAGGGATCAAACCTGCATCCTCGTGGATACTAGTCGGGTTTGTTACCACTAAGTCACAGTGGGAACTCCGGTTCTATTTAAAATTATACATTTATC : 5678
MS : AGGCCAGGGATCAAACCTGCATCCTCGTGGATACTAGTCAGGTTTGTTACCACTAAGTCACAGTGGGAACTCCGGTTCTATTTAAAATTATACATTTATC : 5671
* 5720 * 5740 * 5760 * 5780 * 5800
LW : CTATAAATATAGAAAAGCCTGTATTACACCTATCATCCAAATATGTATAAAATACATCGAAGATAACATATTTATGCATATAGCATTTACGTAAATCTAT : 5778
MS : CTATAAATATAGAAAAGCCTGTATTACACCTATCATCCAAATATGTATAAAATACATCGAAGATAACATATTTATGCATATAGCATTTATGTAAATCTAT : 5771
* 5820 * 5840 * 5860 * 5880 * 5900
LW : ATTTTAAAGAAAAGAAAGAGGGAACTTGTAAACTCAACTAGGCCAGTAGGTTTCACATCTTTTAATTCCAGGAGCCCCTTCTTTGAGCAACAGCTTTCAC : 5878
MS : ATTTTAAAGAAAAGAAAGAGGGAACTTGTAAACTCAACTGGGCCAGTAGGTTTCACATCTTTTAATTCCAGGAGCCCCTTCTTTGAGCAAAAGCTTTCAC : 5871
* 5920 * 5940 * 5960 * 5980 * 6000
LW : AGAGGCCTGATATGAAGCTGCTCTGATTGAAGCCAGCCATGGGAAACCCAAAGGGCTACCGACTGGTTTCCTTCCTGCTTGTCCCCTCCTTGGTGGCTCT : 5978
MS : AGAGGCCCGATATGAAGCTGCTCTGATTGAAGCCAGCCATGGGAAACCCAAAGGGCTACCGACTGGTTTCCTTCCTGCTTGTCCCCTCCTTGGTGGCTCT : 5971
* 6020 * 6040 * 6060 * 6080 * 6100
LW : GAAGGAGCTCTTGGAAACCCC-GGGTAATCGTTTTCAGTGGCAGAAATGTGAAGTGATTACCCTGGGCCATGCAGCCTGCCCGATGACACCTGTTCATTG : 6077
MS : GAAAGAGCTCTTGGAAACCCCCGGGTAATCGTTTTCAGTGGCAGAAATGTGAAGTGATTACCCTGGGCCATGCAGCCTGCCCGATGACACCTGTTCATTG : 6071
* 6120 * 6140 * 6160 * 6180 * 6200
LW : CAATTTCCGTTCCTAGCATATTATTGCCATGTTCCAGAAAAACGACTTGTCTACATAAAAGGCCTTCCCATTCTTGCCTTTGGCACCTTAAAGCAGTAAA : 6177
MS : CAATTTCCGTTCCTAGCATATTATTGCCATGTTCCAGAAAAACGACTTGTCTACATAAAAGGCCTTCCCATTCTTGCCTTTGGCACCTTAAAGCAGTAAA : 6171


* 6220 * 6240 * 6260 * 6280 * 6300
LW : ATATTAGATTTCTGACTGTCCACTGGGGTGCCACTAGACCCATGTGACCAGCGAACACTCAAAATGTAGGCTGTGCCGTGGAGGAACTAAATTTGTTTAA : 6277
MS : ATATTAGATTTCTGACTGTCCACTGGGGTGCCACTAGACCCATGTGACCAGCGAACACTCAAAATGTAGGCTGTGCCGTGGAGGAACTAAATTTGTTTAA : 6271
* 6320 * 6340 * 6360 * 6380 * 6400
LW : ATCAACTTCAATTTAAAGAGCCATATGTTTCGTGGCCACCAAAGCCATGGAGAGTTTCTTCCTTGTTCCATAGAGAAAGAAAACAAGGAAGAGCAGGGCT : 6377
MS : ATCAACTTCAATTTAAAGAACCATATGTTTCGTGGCCACCAAAGCCATGGAGAGTTTCTTCCTTGTTCCATAGAGAAAGAAAACAAGGAAGAGCAGGGCT : 6371
* 6420 * 6440 * 6460 * 6480 * 6500
LW : AAATGAGGGCACCTCTGCTTCAGTTTGGCCTCATCTGAGAGGAATTTCAGCTGCCACACAGAGCACAGGAAAGCA--GTGTTAGCACCAGTGTGAAGAGC : 6475
MS : GAATGAGGGCACCTCTGCTTCAGTTTGGCCTCATCTGAGAGGAATTTCAGCTGCCACACAGAGCACAGGAAAGCACAGTGTTAGCACCAGTGTGAAGAGC : 6471
* 6520 * 6540 * 6560 * 6580 * 6600
LW : CCCTGGGTTTGAACCCCGGCTCTGCCACTGACTCTCACTTGACTTCACTGTGGCTGAATGTCTCCTGATCTGCAAGATGGGATTGAGAATAATGAGGATG : 6575
MS : CCCTGGGTTTGAACCCCGGCTCTGCCACTGACTCTCACTTGACTTCACTGTGGCTGAATGTCTCCTGATCTGCAAAATGGGATTGAGAATAATGAGGATG : 6571
* 6620 * 6640 * 6660 * 6680 * 6700
LW : ACACCTACTCCAGGATTAAGAGAAGGAATATTTATAAATAGTTTTGAAAGTATGCCTGGCTCACAGTCAGTGCTAAGTAAGAGTTTATTAAAGGAATGTT : 6675
MS : ACACCTACTCCAGGATTAAGAGAAGGAATATTTATAAATAGTTTTGAAAGTATGCCTGGCTCACAGTCAGTGCTAAGTAAGAGTTTATTAAAGGAAAGTT : 6671
* 6720 * 6740 * 6760 * 6780 * 6800
LW : CCGGAAAGAAGACCTTATGACAGGTATATCTAAGGCTCTCTTTGACTGTGTCATGCTTGTTATGCTTATCCTAAGTTTTCAGGTTATCACTTAGAGATCC : 6775
MS : CCAGAAAGAAGACCTTATGACAGGTATATCTAAGGCTCTCTTTGACTGTGTCACGCTTGTTATGCTTATCCTAAGTTTTCAGGTTATCACTTAGAGATCC : 6771
* 6820 * 6840 * 6860 * 6880 * 6900
LW : TTGCTTCTTTTTTTCTAGAATGACAAATGTGCAACTCTTTTTTTTTTTTGGCTTTCCAACAAATACACAGATGCTGTAGCCACATTGCTAAAGCCTGACCC : 6875
MS : TTGCTTCTTTTTTTCTAGAATGACAAATGTGCAACTCTTTTTTTTTTTGTAGCTTTCCAACAAATACACAGATGCTGTAGCCACATTGCTAAAGCCTGACCC : 6871
EXON 4
* 6920 * 6940 * 6960 * 6980 * 7000
LW : ATCTCAGAAGCAGACTTTCCTAGCGCCACAGGTACTTTTGATTTTAACTTATTTTAAATATTAAAATTCTCATGGTTAAAGAACCACCACCACAAATTGT : 6975
MS : ATCTCAGAAGCAGACTTTCCTAGCGCCACAGGTACTTTTGATTTTAACTTATTTTAAATATTAAAATTCTCATGGTTAAAGAACCACCACCACAAATTGT : 6971
* 7020 * 7040 * 7060 * 7080 * 7100
LW : AACTACCAAGCAGGTCATTTAGCCTGATGAAATGGCTCTTTCCTGCTGATTGGTAAACATGAACTTCGAGCAACTGATCTGCCATGTGGTCTTCCCTGCA : 7075
MS : AACTACCAAGCAGGTCATTTAGCCTGATGAAATGGCTCTTTCCTGCTGATTGGTAAACATGAACTTCGAGCAACTGATCTGCCATGTTGTCTTCCCTGCA : 7071


* 7120 * 7140 * 7160 * 7180 * 7200
LW : AAGCCATCCACCAATGAGGCTAAAAAAATAA----GACCTGCTGTGGATGGAAAAGGCTACCTTCTGAATTTAAAAATTAAGCATGTATGATGTACACTGA : 7171
MS : AAGCCATCCACCAATGAGGCTAAAAAAATAAATAAGACCTGCTGTGGATGGAACAGGCTACCTTCTGAATTTAAAAATTAAGCATGTATGATGTACACTGA : 7171
* 7220 * 7240 * 7260 * 7280 * 7300
LW : TCTGTGCCCTTCCTTCTTCAGAATACTATTTCCTCGGAGGAAACGGACGACTTCAAACAAGAGGTGAGTTTCACTCTGAATCCAAGGTCCGGCATGTCAC : 7271
MS : TCTGTGCCCTTCCTTCTTCAGAATACTATTTCCTCGGAGGAAACGGACGACTTCAAACAAGAGGTGAGTTTCACTCTGAATCCAAGGTCCGGCATGTCAC : 7271
EXON 5
* 7320 * 7340 * 7360 * 7380 * 7400
LW : AGAGGGACCAAGGAAGCCATTCTTTCTCCTAGTGACACTGCCAGGGCTCAACTCCGGCAAAATTACGTTAGCAAATGATCCAGTAAATATCACGTCTAAA : 7371
MS : AGAGGGACCAAGGAAGCCATTCTTTCTCCTAGTGACACTGCCGGGCTCAACTCCGGCAAAATTACGTTAGCAAATGATCCAGTAAATATCACGTCTAAA : 7371
* 7420 * 7440 * 7460 * 7480 * 7500
LW : ATTTGCTCACCAAAAGAGCAAGTATTTCCTAAATGTCAAAAA-TACCAGGTTTCTACAAATACGTTTACTGTGGCAGGAAGTGTTGACTCATGTCTGTTT : 7470
MS : ATTTGCTCACCAAAAGAGCAAGTATTTCCTAAATGTCAAAAAATACCAGGTTTCTACAAATACGTTTACTGTGGCAGGAAGTGTTGACTCATGTCTGTTT : 7471
* 7520 * 7540 * 7560 * 7580 * 7600
LW : GAAATACTAAGCCTATTGCAGTTATTTAAGGTATGTGGTTGATGTCATCTAATACCTAATAAGTATCTAAATAATGCCTACCAAAAACAAGATTTTGCCT : 7570
MS : GAAATACTAAGCCTATTGCAGTTATTTAAGGTATGTGGTTGATGTCATCTAATACCTAATAAGTATCTAAATAATGCCTACCAAAAACAAGATTTTGCCT : 7571
* 7620 * 7640 * 7660 * 7680 * 7700
LW : CCAGGAACTTTTCTGTTCCTAACATTATAGTTTTTTAATAGCTCTTCTTTTTATTGGGACTATTCATATATTTCCTAAGCAAAGGGATAAGGTAAAGCTG : 7670
MS : CCAGGAACTTTTCTGTTCCTAACATTATAGTTTTTTAATAGCTCTTCTTTTTATTGGGACTATTAATATATTTCCTAAGCAAAGGGATAAGGTAAAGCTG : 7671
* 7720 * 7740 * 7760 * 7780 * 7800
LW : TCTTCCTTAAGATTCCATTTGAAGGTGCTAAGAATCTCTAATAAGGAAAGGTAGATAAATAGCCCAGTATATTTGCTAATCAGATGTTTAATAGACACGA : 7770
MS : TTTTCCTTAAGATTCCATTTGAAGGTGCTAAGAATCTCTAATAAGGAAAGGTAGATAAATAGCCCCGTATATTTGCTAATCAGATGTTTAATAGACACGA : 7771
* 7820 * 7840 * 7860 * 7880 * 7900
LW : ACTGATTTTGAGCTGCAGTATATATATATATATATATAT----------GGATAACAGAGTCACTTTTAATTATCCGCACTAATAGGGAAGGGATTTTGG : 7860
MS : ACTGATTTTGAGCTGCAGTATATATGTATATGTATATATATATATGCACGGATAACAGAGTCACTTTTAATTATCCGCACTAATAGGGAAGGGATTTTGG : 7871
* 7920 * 7940 * 7960 * 7980 * 8000
LW : GTGGGGGGGTACACATTACTGATATATGAATGCTCTGATCGCCTGAGAATTCTCATTTCAAATAGAAAAAGAAATTCCACAGTTAGTATCTGTATTTATT : 7960
MS : GTGTGGGGGGTACACATTACTGATATATGAATGCTCTGATCGCCTGAGAATTCTCATTTCAAATAGAAAAAGAAATTCCACAGTTAGTATCTGTATTTATT : 7971


* 8020 * 8040 * 8060 * 8080 * 8100
LW : ACATTATTGAATTCAGGAAATTGAATAATGTCTAATGTGAACAAGGAAAAACAGTGTTATTACGTACGTATCGTACTCAGATATTTGGCTTGAAAATACC : 8060
MS : ACATTATTGAATTCAGGAAATTGAATAATGTCTAATGTGAACAAGGAAAAACAGTGTTATTACGTACGTATCGTACTTAGATATTTGGCTTGCAAATACC : 8071
* 8120 * 8140 * 8160 * 8180 * 8200
LW : AATGAAAATAATTTTTGTGAGAAAGTCAGAAAAAATGCCTGAATATGATGCCACTTATCATTGTGTTAAAAGAGAAATCAGAAGATGCTTGCTTTAGCGC : 8160
MS : GATTAAAATAATTTTTGTGAGAAAGCCAGAAAAAATGCCTGAATATGATGCCACTTATCATTGTGTTAAAAGAGAAATCAGAAGATGCTCGCTTTAGCGC : 8171
* 8220 * 8240 * 8260 * 8280 * 8300
LW : CTAAGGGCCACGGAGTGCATTGTCCTGACTGCAAAGATTCTCACACGAAAGGAAGATAGTAAGCGGCAGATGCACTGTCGAGCTAAGGGGAAATTAAAAA : 8260
MS : CTAAGGGCCACGGAGTGCATTGTCCTGACTGCAAAGATTCTCACACGAAAGGAAGATAGTAAGCGGCAGGTGCACTGTCGAGCTAAGGGGAAATTAAAAA : 8271
* 8320 * 8340 * 8360 * 8380 * 8400
LW : CTCTCACACCCATGTTTCTGGCCATCTCGATATTTAGACCCTGCCAAGCAAGTCCAACGAAAGCCCTGAGCAAACAGACGATGTGGACGACGACGACGAC : 8360
MS : CTCTCACACCCATGTTTCTGGCCATCTCGATATTTAGACCCTGCCAAGCAAGTCCAACGAAAGCCCTGAGCAAACAGACGATGTGGACGACGACGACGAC : 8371
EXON 6
* 8420 * 8440 * 8460 * 8480 * 8500
LW : GAAGACCACGTGGACAGCAGGGACACGGACTCCGAGGAAGCTGATCACCCTGACGACGCTGACCGATCCGACGAGTCTCATCACTCCGATGAATCCGATG : 8460
MS : GAAGACCACGTGGACAGCAGGGACACGGACTCCGAGGAAGCTGATCACACTGACGACGCTGACCGATCCGACGAGTCTCATCACTCCGATGAATCCGATG : 8471
* 8520 * 8540 * 8560 * 8580 * 8600
LW : AGCTGGTCACCGATTTCCCCACCGACACCCCAGCAACCGACGTCACTCCGGCTGTCCCCACCGGAGACCCCAATGATGGCCGCGGGGATAGTGTGGNNTA : 8560
MS : AGCTGGTCACCGATTTCCCCACCGACACCCCAGCAACCGACGTCACTCCGGCTGTCCCCACCGGAGACCCCAATGATGGCCGCGGGGATAGTGTGGCCTA : 8571
* 8620 * 8640 * 8660 * 8680 * 8700
LW : TGGACTGAGGTCAAAATCTAAGAAGTTCCGCAGATCCGAAGCCCAGGTAAATCCTGAAACAGACGCAGCCGATGGTTCCGAGGAGAGCCCTGTCCTAGGA : 8660
MS : TGGACTGAGGTCAAAATCTAAGAAGTTCCGCAGATCCGAAGCCCAGGTAAATCCTGAAACAGACGCAGCCGATGGTTCCGAGGAGAGCCCTGTCCTAGGA : 8671
* 8720 * 8740 * 8760 * 8780 * 8800
LW : AACCAGAATCGGCAGGCTCATGTATTGGCTCATTCAGCCAGCATGACTACTTGAACACAAAACCTGTGTTTAAATCCACATAATTTCTCCCTAATTTTGT : 8760
MS : AACCAGAATCGGCATGCTCATGTATTTGCTCATTCAGCCAGCATGACTACTTGAACACAAAACCTGTGTTTAAATCCACATAATTTCTCCCTAATTTTGT : 8771
* 8820 * 8840 * 8860 * 8880 * 8900
LW : CTCTCAATCACGAGGCAGTTTTTCCAAACCCTGGCTATAAAGCACTTATTCTATTCACTGTTTTAAATTTAAAAAGGACTTCCGAAAAGAGTAATTCTTT : 8860
MS : CTCTCAATCACGAGGCAGTTTTTCCAAACCCTGGCTATAAAGCACTTATTCTATTCACTGTTTTAAATTTAAAAAGGACTTCCGAAAAGAGTAATTCTTT : 8871


* 8920 * 8940 * 8960 * 8980 * 9000
LW : T CATGCTATCATGGTCAC TTTTTGAAATTCAATGCAGAAGA AGA CTTGGAGAGGTGCAAATTACACTGACTCCACAGTCAAAGGATCTGAATTTGGCAC : 8960
MS : T CATGCTATCATGGTCAC TTTTTGAAATTCAATGCAGAAGA --- CTTGGAGAGGTGCAAATTACACTGACTCCACAGTCAAAGGATCTGAATTTGGCAC : 8968
* 9020 * 9040 * 9060 * 9080 * 9100
LW : ATGGCTTTGCCATTTACTACTCACATTACTTTTTAACTTCT TTTTTT


* 9820 * 9840 * 9860 * 9880 * 9900
LW : GAGTCAGCCGGATGACCGCAGTGTGGAAACCCGCAGCCAGGAGCAGTCCAAAGAATACACGATCAAGACCTATGATGGGAGCAATGAGCATTCCAATGTG : 9860
MS : GAGTCAGCCGGATGACCGCAGTGTGGAAACCCGCAGCCAGGAGCAGTCCAAAGAATACACGATCAAGACCTATGATGGGAGCAATGAGCATTCCAATGTG : 9525
* 9920 * 9940 * 9960 * 9980 * 10000
LW : ATTGAGAGTCAGGAAAATTCCAAAGTCAGCCAAGAATTCCACAGCCATGAAGACAAGCTGGTCCCAGACTCTAAGAGCGAAGAAGACAAACACCTGAAAC : 9960
MS : ATTGAGAGTCAGGAAAATTCCAAAGTCAGCCAAGAATTCCACAGCCATGAAGACAAGCTGGTCCCAGACTCTAAGAGCGAAGAAGACAAACACCTGAAAC : 9625
* 10020 * 10040 * 10060 * 10080 * 10100
LW : TTCGAGTTTCTCATGAATTAGAGAGTGCGTCTTCTGAGATCAACTGAGAGAAATACAATGTCTTACTTTGCTTTTAGTAAAAAGAAAAAAAAACACATTG : 10060
MS : TTCGAGTTTCTCATGAATTAGAGAGTGCGTCTTCTGAGATCAACTGAGAGAAATACAGTGTCTTACTTTGCTTTTAGTAAAAAGAAAAAAAAATACATTG : 9725
Stop
LW : TAGC : 10064
MS : TAGC : 9729

Appendix III

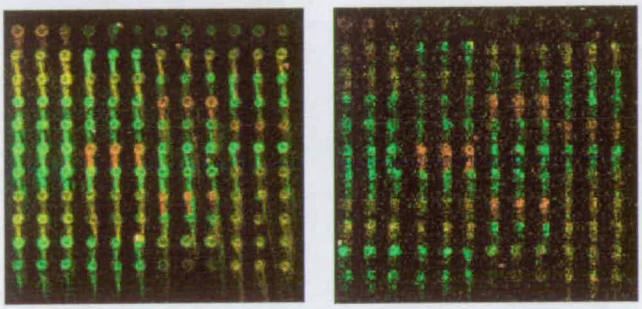
Experiments used to test the methodology of producing and hybridising microarray slides:

i) **Hypothesis:** The comet tails will be removed by pre-blocking the slides. Blocking the slides will also reduce the intensity of the background noise.

Methods: Before denaturing the slides, immerse them in BSA solution (1% w/v Bovine Serum Albumin V (Sigma), 3x SSC and 0.1% SDS) and block them for 20 minutes in a water bath at 55° C.

Results: An example of an array on replicate slides, where the formation of comet tails was particularly noticeable without BSA block.

Without BSA block **With BSA block**



With the BSA block the comet tails do seem to be reduced, however so does the spot intensity. When these two slides were analysed there was too much overlap on the scatter plot between the three classes of genotypes to be able to differentiate between them, however the results were clearer for the slide without the block. This indicates that the background level was too high for both slides. It was realised that by ensuring that when denaturing the slides the water was boiling vigorously and the slides were immersed for no less than 60 seconds, the formation of comet tails could be greatly reduced. Therefore blocking of the slides would not be performed, in order to maximise the intensity of the signal.

ii) **Hypothesis:** The 48-pin tool does not print as clearly as the 16-pin tool.

Methods: Slides were reprinted using the 16-pin tool and two 4x4 array grids were printed on each slide.

Results: The spots on the re-printed slides had a "do-nut" morphology, suggesting that the DNA spots had dried out too much either during the printing

of the slides or possibly during baking. It was therefore not possible to analyse the intensity of the spots and compare the quality of the slides with those previously printed with the 48-pin tool.

iii) **Hypothesis:** Adding a higher concentration of probe will increase the fluorescence intensity of the spot.

Methods: Hybridise slides from the same batch with 6 μM, 5 μM and 3 μM of each probe (previously used 3 μM).

Results: The spots on the slides probed with 6 and 5 μM were fainter than with 3 μM. It appears that 3 μM is the optimum concentration of probe, any higher than this results in less efficient binding of the DNA to the "target DNA" on the slide.

iv) **Hypothesis:** Less stringent wash conditions will increase the intensity of the spots.

Methods: The slides were washed in 5x SSC with 0.05 % SDS for both 10 and 5 minutes and then 5x SSC for both 10 and 5 minutes. The final wash in water was removed and the slides were spun immediately to prevent the salt from the SSC wash drying onto the slide.

Results: The spots appeared to be brighter than the original slides and the slides washed for 5 minutes were brighter than those washed for 10 minutes. However for both, the intensity of the background was also increased and therefore the net result was no better than before.

v) **Hypothesis:** The probes are light sensitive and repeated freeze thawing exposes them to the light and bleaches them. Therefore by making several aliquots from the stock solution and using each one only once for a single hybridisation reduces the loss of their activity.

Methods: Hybridise slides with fresh aliquots of probes at a concentration of 3 μM.

Results: The spot intensity was higher than with the original experiment and due to the more vigorous denaturing conditions no comet tails were observed.

Freeze/thawed probes **Fresh aliquots of probes**